REPORT RESUMES

ED 010 627 24
IDENTIFICATION AND DEFINITION OF SUBJECT-MATTER CONTENT
VARIABLES RELATED TO HUMAN APTITUDES, VOLUME 1.
BY- KROPP, RUSSELL P. AND OTHERS
FLORIDA ST. UNIV., TALLAHASSEE, INST. OF HUMAN LNG
REPORT NUMBER CRP-2914-1 PUB DATE JAN 67
REPORT NUMBER BR-5-0326-1
CONTRACT OEC-5-10-297
EDRS PRICE MF-$0.36 HC-$9.40 235P.

DESCRIPTORS- *INSTRUCTIONAL DESIGN, INSTRUCTIONAL IMPROVEMENT, *APTITUDE, COGNITIVE ABILITY, *INSTRUCTIONAL MATERIALS, TEACHING METHODS, LEARNING DIFFICULTIES, *LEARNING THEORIES, *INDIVIDUALIZED PROGRAMS, ACADEMIC ACHIEVEMENT, MATERIAL DEVELOPMENT, REDUNDANCY, SET THEORY, CHEMISTRY INSTRUCTION, CONCEPT TEACHING, VOCABULARY DEVELOPMENT, MATHEMATICS INSTRUCTION, AUDIOVISUAL AIDS, TALLAHASSEE, FLORIDA

AN EXPLORATORY STUDY WAS CONDUCTED ON THE INTERACTION BETWEEN COGNITIVE APTITUDES AND VARIED INSTRUCTIONAL TREATMENTS (TEACHING MATERIALS AND TEACHER METHODS) TO DETERMINE IF LEARNING DIFFICULTIES COULD BE SIGNIFICANTLY MINIMIZED BY ALTERING THE CONTENT OF INSTRUCTIONAL TREATMENTS TO FIT THE COGNITIVE APTITUDE PATTERNS OF INDIVIDUAL LEARNERS. THE AUTHORS SUGGESTED THAT ACHIEVEMENT OF THE STUDENT WOULD BE DIRECTLY RELATED TO THE CONGRUENCE BETWEEN HIS PATTERN OF APTITUDES AND THE "FORM OF CONTENT" OF THE MATERIAL TO BE LEARNED. TO SUBSTANTIATE THIS CONTENTION, A PROGRAM OF HIGHLY SIMPLIFIED STUDIES WAS STRUCTURED AND CONDUCTED. SEVEN SERIES OF STUDIES WERE UNDERTAKEN AMONG DIFFERENT GRADE LEVELS--(1) "FORM OF CONTENT" VARIABLES IN EXISTING TEXTUAL MATERIALS WERE IDENTIFIED, (2) REDUNDANCY AS A "FORM OF CONTENT" VARIABLE WAS EXAMINED, INCLUDING REDUNDANCY LEVELS AND COGNITIVE APTITUDES RELATED TO REDUNDANCY, (3) MATERIALS ON SET THEORY CONCEPTS WERE DEVELOPED AND TAUGHT, EMPHASIZING THE APTITUDE COMBINATIONS OF VERBAL-DEDUCTIVE, VERBAL-INDUCTIVE, FIGURAL-DEDUCTIVE, AND FIGURAL-INDUCTIVE, (4) ACHIEVEMENT OF STUDENTS IN CHEMISTRY WAS RELATED TO SIMILARITIES OF TEACHER AND STUDENT APTITUDE PATTERNS WHEN GENERAL ABILITY WAS HELD CONSTANT, (5) PERFORMANCE ON VERBAL AND FIGURAL CONCEPT ATTAINMENT TASKS, PRESENTED BY EITHER COMPUTER-ASSISTED INSTRUCTION OR BY A HUMAN EXPERIMENTOR, WAS RELATED TO APTITUDE VARIABLES BELIEVED TO BE RELEVANT TO THE TASKS, (6) INTERACTIONS OF APTITUDE AND INSTRUCTION IN VOCABULARY LEARNING WERE STUDIED, AND (7) THE EFFECTS OF SYMBOLIC AND SEMANTIC CONTENT ON THE LEARNING OF MATHEMATICAL OPERATIONS WERE ANALYZED. RESULTS OF THESE STUDIES DEMONSTRATED THE FEASIBILITY OF CONTROLLING APTITUDE-INSTRUCTION INTERACTIONS TO ENHANCE LEARNING IN A VARIETY OF SUBJECT-MATTER AREAS AND AT DIFFERENT AGE AND GRADE LEVELS. RELATED INFORMATION MAY BE FOUND IN ED 010 628. (JH)

ED010627

5-0326

FINAL REPORT
Project No. 2914
Contract No. OE-5-10-297

U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
Office of Education
This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated do not necessarily represent official Office of Education position or policy.

IDENTIFICATION AND DEFINITION OF SUBJECT-MATTER CONTENT VARIABLES RELATED TO HUMAN APTITUDES

Volume I

January 1967

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Bureau of Research

IDENTIFICATION AND DEFINITION OF SUBJECT-MATTER CONTENT VARIABLES RELATED TO HUMAN APTITUDES

Project No. 2914
Contract No. OE-5-10-297

Investigators

Russell P. Kropp
Willard H. Nelson
F. J. King

Faculty Participants	Staff
W. Louis Bashaw	John B. Davis
Jacob G. Beard	Dennis M. Roberts

January, 1967

The research reported herein was performed pursuant to a contract with the Office of Education, U.S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

Institute of Human Learning
Florida State University
Tallahassee, Florida

PREFACE

The contents of this report are the results of exploratory studies of aptitude treatment interactions as they might occur in public school instructional situations.

About three years ago several of us began mutual discussions of the possibility that different cognitive aptitudes might be elicited by different sets of instructional materials and that these aptitudes might also be differentially elicited by different instructional methods. To some extent these opinions arose from comments by undergraduate students who, when receiving academic counseling, would comment "Professor Jones teaches the way I learn, so I want to get in his section again." Also, these opinions were based on the implicit implications of perplexingly different comments by students who were experiencing difficulty in class and would express their difficulty as "I can't see it," "I don't get the feel of it," which seem to suggest a preferred mode of gaining understanding. These opinions were also supported by observations by instructors who used different approaches in an elementary statistic course, e.g., logical development and geometric explanations, when teaching bivariate correlation. They observed that only some students profited from a particular explanation whereas others profited from an alternate explanation of the concept. Finally, our examination of graded textbooks for the elementary school led us to believe that items in the series were relatively constant with regard to objectives but differed with regard to the degree of verbal aptitude which was needed to cope with them.

The joint effect of these opinions and observations caused us to infer tentatively that different instructional materials and teaching methods--separately and, quite likely, jointly--cater to some learners and not others on the basis of the aptitudes which these require for learning from them. The observation about graded textbooks led inevitably to wondering whether instructional materials could be written which would differ from each other with respect to the several aptitudes which they would require in order to profit from them, instead of differing almost solely on verbal load.

This rather loose rationale was discussed by us and our view of it matured so that it took this form: Learning, at least in part, is a result of the interaction of learner aptitudes and variables which are unwittingly embedded in or unavoidably part of textual materials and teaching behavior. Further refinement led to this proposition: Learning will be optimized when instructional materials and teaching behavior call into the learning situation the most highly developed aptitudes of the students.

Our interest in this notion arose from or was stimulated by certain beliefs and facts. First, if such a conception were empirically valid, then it would be a potentially powerful theory with which to guide the educational diagnosis of learning difficulties. Second, it would provide at least a theoretical basis on which to hold that any student is potentially capable of learning any content, a view held by many today but without noticeable empirical or theoretical support. Third, it enabled us to understand the student comments about their preference for instructors and their learning difficulties. Fourth, because all of us were in some manner involved in research on computer-assisted instruction, it appeared to be the first-step toward that remote point at which a computer might be programmed to prepare a unique set of instructional materials for each student on the basis of his pattern of aptitudes.

Planning an empirical investigation of the general hypothesis confronted us with the need to make decisions about certain critical questions as a prelude to structuring any program of studies at all. These decisions were made for good or ill and determined the particular kinds of studies which were undertaken. Some of these choice-points and the decision made about each of them are discussed here simply because other researchers might wish to make contrary choices so that eventually the results of all studies might be complementary. First, aptitude has several meanings and falls in several domains of behavior. We choose to regard aptitude as a learned behavior which explains transfer in learning situations. Also, we narrowed our interest to cognitive aptitude, although recognizing the existence and potential importance of psychomotor and attitudinal aptitudes.

This latter choice was based on personal preference of the investigators, their belief that our understanding of cognitive aptitude is more advanced than is our grasp of other aptitudes, the availability of more instruments for assessing cognitive aptitudes than other aptitudes, and that cognitive aptitudes have greater immediate relevance from the standpoint of their amenability to manipulation than do the others. Second, we recognized that the study of this topic would be long-term and that several entry points were feasible when judged in terms of potential educational significance. One might launch studies to alter experimentally the aptitude levels of students, thereby shaping the students to "fit" the prevailing of possibly other more economical teaching methods and materials. One might launch frontal studies of the interaction of aptitude and instruction. We chose the latter but continued to be divided among ourselves about which avenue is more important. Third, we firmly decided at the outset to conduct studies which had educational "face-validity;" i.e., we wished to avoid laboratory-studies the conditions and results of which would be difficult to replicate in or generalize to practical educational situations. The reader will eventually note that the complexity and variety of pertinent variables eventually drove us into the laboratory despite our wish to avoid it. Fourth, we decided to capitalize on existing textual materials and teaching methods. The former would be analyzed for hitherto unidentified variables, which we called form of content variables, and the latter would be subjected to analyses which were similar in purpose. The reader will note subsequently that this decision had to be abandoned because of the unsuspected complexity of both variables (or, perhaps, the conspicuous limitations of the investigators) and our logistical inability to study them in natural situations. Consequently, the majority of studies reported here are highly simplified and usually are based on materials which we specifically constructed for the purpose.

* * * * * * * *

A number of persons made contributions to this project in addition to the investigators, faculty participants, and staff who were listed on the title page. Their assistance is gratefully acknowledged here.

Project graduate assistants. Mr. Donald L. Coan, graduate student in the Department of Guidance and Counseling,

assisted in test scoring and analysis. Mr. John W. Korp, now a high school teacher, assisted in the preparation of learning materials and test administration. Mr. Lloyd Lewis, graduate student in the Department of Educational Research and Testing, assisted in the development of a computer assisted instruction program. Mr. Donald W. Shontz, graduate student in the Department of Educational Research and Testing, assisted in data collection and analysis. Mr. Carl Neil Shaw, graduate student in the Department of Educational Research and Testing, assisted in test selection and data analysis.

Non-Project graduate students. Mr. Aaron Bauldree, graduate student in the Department of Educational Research and Testing, assisted in developing and in collecting data with a computer assisted instruction program. Dr. Bruce Clear, now at Florida Agricultural and Mechanical University, conducted in collaboration with project personnel one of the studies reported herein. Dr. Emmett Kohler, now at the University of Georgia, conducted in collaboration with project personnel one of the studies reported herein. Mr. William Lukes, now of the Armed Forces, assisted in data processing.

Consultants. Dr. Lee J. Cronbach and Dr. Richard E. Snow, both of Stanford University, served as general consultants. Dr. Ralph Hoepfner, University of Southern California, served as general consultant and generously provided many tests of the Structure of Intellect model developed by the Aptitudes Research Project.

* * * * * * * *

We wish to acknowledge our appreciation to the Educational Testing Service for permitting us to reproduce tests which appear in the Kit of Reference Tests.

A considerable amount of field data was collected during the term of this project and it was made possible by the unstinting and unfailing cooperation of the administrators and instruction personnel of the following educational units: Broward County (Ft. Lauderdale, Florida) School System; Dade County (Miami, Florida) School System; Escambia County (Pensacola, Florida) School System; Florida State University School; Gadsden County (Quincy, Florida) School System; Hillsborough County (Tampa, Florida) School System; Leon County

(Tallahassee, Florida) School System; Sarasota County (Sarasota, Florida) School System; Department of Mathematics and Psychology of East Carolina College, Greenville, North Carolina; and Departments of English Education, Psychology and Science Education of Florida State University.

We wish also to express our appreciation to the staff of the Florida State University Computing Center and to the National Science Foundation for its support of the Computing Center.

Finally, we wish to express our gratitude to the Bureau of Research of the U.S. Office of Education for its support of the project. Without it, the project could not have been undertaken. Whatever good might emanate from the results which appear in this report must be attributed in large measure to the U.S. Office of Education.

TABLE OF CONTENTS

VOLUME I

Page

PREFACE. ii

Chapter

I. INTRODUCTION. 1

II. REVIEW OF LITERATURE. 4

A. Theoretical Papers. 4
B. Empirical Papers. 8

III. FORM OF CONTENT VARIABLES IN TEXTUAL MATERIALS. 14

A. Identification of Variables; Analysis of Single Passages. 14
B. Identification of Variables; by Comparison of Equivalent Passages. . . 19
C. Summary. 25

IV. REDUNDANCY AS A FORM OF CONTENT VARIABLE. . 27

A. Redundancy, Content Area and Expressional Style. 28
B. The Relation of Cloze Scores to Selected Cognitive Variables. 42
C. Relative Redundancy of Reading Material with Different Readability Levels and Parallel Content. 53
D. Reading Comprehension as a Function of the Interaction of Learner Aptitude and Redundancy. 57
E. Summary. 67

V. ELEMENTARY SET CONCEPTS. 69

A. Introduction. 69

Chapter Page

B. Method. 69
C. Results 72
D. Discussion. 86
E. Summary. 86

VI. STUDENT AND TEACHER ABILITY PATTERNS AND ACHIEVEMENT IN CHEM-STUDY. 88

A. Introduction. 88
B. Method. 89
C. Results. 91
D. Discussion. 94
E. Summary. 94

VII. CONCEPT ATTAINMENT FROM VERBAL AND FIGURAL INSTANCES. 98

A. Introduction. 98
B. Method. 99
C. Results. 107
D. Discussion. 111
E. Summary. 111

VIII. VOCABULARY LEARNING. 113

A. Identification of Relevant Abilities for Two Methods of Learning Vocabulary Words. 114
B. Pilot Investigation of Three Methods of Learning Vocabulary Words. 119
C. An Investigation of Three Methods of Learning Vocabulary Words. 127
D. Two Methods of Vocabulary Learning at Different Stages of Practice. 135
E. Summary. 150

IX. MATHEMATICAL OPERATIONS. 152

A. Symbolic and Semantic Learning of Mathematical Operations. 153
B. Symbolic and Semantic Learning of Mathematical Operations: Replication with a Similar Population. 167
C. Symbolic and Semantic Learning of Mathematical Operations: Replication with a Dissimilar Population. 185
D. Summary. 202

Chapter Page

X. DISCUSSION. 204

   A. Interaction of Treatment-Ability-Criterion. 204
   B. Treatment Duration. 205
   C. Reliability of Differences. 206
   D. Time as a Criterion. 207
   E. Identification of Relevant Abilities. . 208
   F. Need for Better Aptitude Measures. 208
   G. Representativeness of Samples. 209
   H. Generality over Age and Grade... 210

XI. CONCLUSIONS, IMPLICATIONS AND RECOMMENDATIONS. 211

   A. Conclusions. 211
   B. Implications and Recommendations. . . . 212

XII. SUMMARY. 214

XIII. REFERENCES. 219

I. INTRODUCTION

A considerable body of knowledge about differences in human abilities has been produced by differential psychologists and psychometricians. Experimental psychologists conducted studies that led to formulating rough outlines of learning theories based on the average performance of groups of subjects. The readiness of experimentalists to consider the performance of individual subjects has contributed to the current interest in merging these two fields, individual differences and learning theory. In a recent symposium (26), several psychologists addressed themselves to the task of accomplishing such a merger. Consideration of the parameters of learning, as they apply to individual patterns of abilities, will surely influence the development of curriculum materials and teaching methods in a manner that will cause them to be more efficient than those now in use.

In a paper, "How Can Instruction Be Adapted to Individual Differences?", Cronbach (13) discussed how instruction has been adapted to individual differences. He suggested that the most adequate way of coping with individual differences in the school might be to alter instructional methods to fit the aptitude pattern of the learner. This procedure would not be useful unless an aptitude treatment interaction exists; i.e., students high in a given ability achieve better under one method of instruction than they do under another. Ideally, there should be a general theory that would specify the conditions of instruction which are optimal for students of certain ability types so that adapting instructional methods to individual differences can be accomplished systematically.

The general purpose of the research reported herein was to be achieved by completing a three-part sequence of activities: (a) to identify and define operationally dimensions of an hypothesized variable, tentatively called "form of content," which is described subsequently and is believed to be present in all textual materials; (b) to prepare several sets of instructional materials, each member of each set having an unique and known value on a dimension of the form of content variable and all members of a set leading to achievement of the same behavioral outcome; and (c) to determine whether students who use these materials achieve mastery levels that are predictable on the basis of the congruence of the particular value of the form of content variable and the students' cognitive aptitudes.

The importance of the problem arises from the centrality of the form of content variable to the following conception of the teaching-learning process. The last activity (c, above) is based upon the postulate that form of content is a mediating variable in the learning process. Its interaction with the cognitive aptitudes which the student brings to the learning situation determines his attained level of achievement. It is hypothesized that each variation of form of content can be described in terms of the cognitive aptitudes that are associated with maximal achievement of the content in which it is embedded. Consequently, it is hypothesized that student achievement will be directly related to the congruence between his pattern of aptitudes and the form of content of the material to be learned.

Form of content undoubtedly has many dimensions or facets and concerted study of it will probably cause it to be partitioned into several variables. A superficial consideration of instructional materials reveals features, each of which probably facilitates or impedes learning, depending on particular aptitudes of the student. For example, most textual materials place great stress on verbal facility; however, some of these might be recast so that the verbal loading is minimized or another feature (hopefully related to an aptitude) is emphasized. Thus, it is conceivable to have two geometry textbooks, one of which places a premium on verbal ability and the other on spatial ability. The development of topics in some textual materials proceeds inductively, whereas in others it progresses deductively. Both kinds of reasoning have been recognized as cognitive abilities and measured for some time by psychological test specialists. Instructional materials in history might be organized chronologically or topically. Little is known about the differential appeal of such variant organizations of content. Another dimension might be the particular mode of presenting a concept, e.g., does the author of an elementary statistics book when developing the concept of correlation do so by illustrating geometrically that it is the cosine of two angles formed by test vectors, or by illustrating that it is ratio of shared variation to total variation where each is represented by symbols describing operations on scores, or by illustrating nonmathematically that it is the proportion of behavioral elements common to both measures.

The remainder of Volume I of the report is organized as follows: First a review of pertinent literature is presented. The next seven sections contain studies which were done during the project. The last three sections contain a general discussion of problems encountered in the research studies; conclusions, recommendations, and implications; and a general summary. Volume II consists of appendices which contain all of the experimental versions of instructional materials which were produced during the project. These materials represent a substantial investment of time and energy and are included for whatever advantage they might give to other researchers.

II. REVIEW OF LITERATURE

The purpose of this section is to present a review of selected papers and studies that deal directly, theoretically or empirically, with aptitude treatment interactions (ATI), and that are directly applicable to the studies executed in the present project. Theoretical papers are reviewed first and then empirical studies are reviewed.

A. Theoretical Papers

Guilford and Hoepfner

Guilford's Structure of Intellect model (28) is basic to much of the research in this report. In general, the model can be described as a representation of an information processing system having several dimensions. In its present form, the model is three-dimensional; the dimensions being "operations," "content," and "products." The operation dimension is similar to the more traditional constructs of mental acts or thoughts. The content dimension is related to observable stimulus elements. The product dimension seems to be best described as particular kinds of observable responses. The operation, content, and product dimensions respectively contain five, four and six categories each. The definition of each dimension and its categories follows:

OPERATIONS

Major kinds of intellectual activities or processes; things that the organism does with the raw materials of information, information being defined as "that which the organism discriminates."

Cognition. Immediate discovery, awareness, rediscovery, or recognition of information in various forms; comprehension or understanding.

Memory. Retention or storage, with some degree of availability, of information in the same form it was committed to storage and in response to the same cues in connection with which it was learned.

Divergent Production. Generation of information from given information, where the emphasis is upon variety and quantity of output from the same source. Likely to involve what has been called transfer. This operation is most clearly involved in aptitudes of creative potential.

Convergent Production. Generation of information from given information, where the emphasis is upon achieving unique or conventionally accepted best outcomes. It is likely the given (cue) information fully determines the response.

Evaluation. Reaching decisions or making judgments concerning criterion satisfaction (correctness, suitability, adequacy, desirability, etc.) of information.

CONTENTS

Broad classes or types of information discriminable by the organism.

Figural. Information in concrete form, as perceived or as recalled possibly in the form of images. The term "figural" minimally implies figure-ground perceptual organization. Visual spatial information is figural. Different sense modalities may be involved, e.g., visual, kinesthetic.

Symbolic. Information in the form of denotative signs, having no significance in and of themselves, such as letters, numbers, musical notations, codes, and words, when meanings and form are not considered.

Semantic. Information in the form of meanings to which words commonly become attached, hence most notable in verbal thinking and in verbal communication but not identical with words. Meaningful pictures also often convey semantic information.

Behavioral. Information, essentially non-verbal, involved in human interactions where the attitudes, needs, desires, moods, intentions, perceptions, thoughts, etc., of other people and of ourselves are involved.

PRODUCTS

Forms that information takes in the organism's processing of it.

Units. Relatively segregated or circumscribed items of information having "thing" character. May be close to Gestalt psychology's "figure on a ground."

Classes. Conceptions underlying sets of items of information grouped by virtue of their common properties.

Relations. Connections between items of information based upon variables or points of contact that apply to them. Relational connections are more meaningful and definable than implications.

Systems. Organized or structured aggregates of items of information; complexes of interrelated or interacting parts.

Transformations. Changes of various kinds (redefinition, shifts, or modification) of existing information or in its function.

Implications. Extrapolations of information, in the form of expectancies, predictions, known or suspected antecedents, concomitants, or consequences. The connection between the given information and that extrapolated is more general and less definable than a relational connection (28, pp. 3-4).

Ferguson

Ferguson (16, 17) attempted to include human learning and human ability in a single theory. Ability was defined as the performance of an individual in a specified situation whereas learning refers to changes, with repetition, in ability to perform a specified task. Ferguson suggested that conventional learning curves are simply descriptions of changes in abilities with repetition. Abilities reach crude stable limits in the adult population and are therefore considered to be overlearned acquisitions. Transfer of training is considered to be the general case with learning being a special aspect of it. Ferguson pointed out, in hypothesis form, that abilities exert their effects differentially in any learning situation; that different abilities exert different effects at different stages of learning and that the abilities which transfer and produce their effect at one stage of learning might be different from those which transfer and produce their effects at another stage.

Gagné

The two quotations from Gagné which appear below deal with ATI in a specific instructional area.

> . . . it should be possible to verify a number of hypotheses, not only about the nature of abilities involved in mathematics but also about the essential nature of what is learned in mathematics, such as:
>
> 1. People who are high in spatial ability should acquire mediating spatial concepts more readily than they do symbolic or verbal ones. This should also be true for numerical ability and verbal ability.
> 2. People who are high in verbal ability should reveal a better performance on geometric problems when they are taught verbal mediators than when they are taught spatial or symbolic ones. Similar statements could be made for the other basic abilities, spatial and numerical, using algebraic and formal logic problems as performance criteria.
> 3. When people high in spatial ability are taught the same mathematical principle in terms of spatial, verbal, and symbolic concepts, measuring retention of this principle after a period of time should reveal increasing dependence on spatial concepts. Again, similar hypotheses could be made to pertain to the other abilities (23, p. 53).

Gagné also proposed a research design for the study of a specific problem. His statement of the "problem" encapsules the general concerns of the present investigators.

> Differences in fundamental abilities appear to be prominent in the learning of mathematics, as well as in the way people use mathematical concepts. Well-established factors in human abilities are spatial, numerical, and verbal. Although there are studies which have revealed moderate to high correlations between aptitude measures and grades in mathematics, no studies have been conducted in the attempt to make specific predictions concerning the facilitation of different kinds of conceptual learning by different fundamental abilities. The possession of a high degree of spatial ability should facilitate the learning of spatial concepts; high verbal ability should facilitate the learning of verbal concepts; and high numerical ability should facilitate the learning of symbolic concepts.
>
> The learning of concepts of addition of directed numbers may be done verbally, spatially, or symbolically. Verbal rules are perhaps the best known method, occurring in most conventional textbooks. Spatial concepts have been used with considerable success, notably in the textbook of the University of Illinois Committee on School Mathematics. Symbolic concepts can readily be designed to serve the same purpose; in one form, they might resemble some of the symbolism of boolean algebra. Thus in this mathematical topic, the opportunity exists of relating differences in fundamental abilities tc ease of learning the different types of concepts, as well as to final performance in problem solving (24, p. 112).

The references cited above, together with the work of Cronbach referred to earlier, provide strong support for the present series of studies. The conceptual framework provided by the structure of intellect for the classifications of both individual aptitudes and task variables was of much value in planning and executing several studies. The general theoretical formulations of Ferguson and the specific concerns of Gagné aided in the formation of the investigators' ideas about ATI and gave support to their conviction that the area was an important one to investigate.

B. Empirical Papers

Empirical studies which show some support for ATI are presented next.

Dunham, Guilford, and Hoepfner

Dunham, Guilford, and Hoepfner (14) investigated the relationships of abilities of the "class" by "content cells" of the SI model and concept learning. Aptitude tests measuring fifteen of Guilford's factors were administered to a group of 177 junior and senior high school students. Three concept-learning tasks utilizing the same type of content (symbolic, figural, and semantic) were given to all students. Ninety-six examples were given for each of the three types of concept problems and were administered in twelve trials (practice stages) of eight examples each.

The two principal dependent variables were the number of correct responses per trial and the number of correctly verbalized concepts for each task. The results (in terms of factor loadings greater than .30) indicated that the divergent production of symbolic classes and the memory of symbolic class factors were significantly related to certain practice stages in the symbolic-concept learning task. Memory of symbolic classes was significantly related to performances on the figural concept task, and the memory of semantic classes, memory of symbolic classes and convergent production of semantic class factors were important in the semantic-concept learning task. Also, they found that the factor of cognition of semantic classes was strongly related to the verbalization scores in the semantic and figural concept learning tasks.

Gagné and Paradise

Gagné and Paradise (25) reported a study of the interactions of cognitive variables with achievement when teaching strategy was held constant. Their purpose was to investigate a theory that knowledge of a particular content can be divided into an hierarchical series of learning sets. The theory postulates that attainment of any given learning set is dependent on recall ability of subordinate learning sets, instructions defining a new task, and the integration by the learner of subordinate learning sets into the solution of the new task. The authors suggested that any learning task can be analyzed into subordinate learning tasks. These subordinate tasks can be analyzed in turn. The analytic procedure can be continued until a basic set of necessary cognitive abilities is defined as the lowest level of the hierarchy.

They analyzed an algebra task (solving equations) into an hierarchical structure of learning tasks. They determined that the basic cognitive abilities necessary to learn the task were number (arithmetic operations), integration (following directions), and symbol recognition (picture-number associative memory). Measures of these abilities were obtained on 118 seventh-graders by administering one test for each ability. The subjects were taught the algebra task by a programmed textbook. Analyses of the data revealed that the correlations between the three relevant abilities--number, integration, and symbol recognition--and final performance were higher than were correlations between final performance and abilities judged to be irrelevant (vocabulary and speed of symbol discrimination). Correlations between the three relevant abilities and rate of attainment decreased progressively as learning progressed through the hierarchy of learning sets, whereas correlations between irrelevant variables and rate of attainment remained nearly constant. Correlations between number and symbol recognition and achievement of learning sets increased as level increased, whereas correlations between irrelevant variables and achievement of learning sets did not.

The results of the Gagné and Paradise study present a strong argument for the thesis that a particular set of abilities is important in a given learning task. The investigators did not discuss the problem of teaching students who were weak in any of the three basic abilities.

Blade and Watson

Blade and Watson (4) reported that engineering students increased their scores on the CEEB Spatial Relations Test by nearly one standard deviation and nonengineering students increased by only one-half standard deviation during their freshman year in college. Engineering students in one institution (N = 89) were studied intensively with regard to these gains. Previous courses in mechanical drawing, mechanical hobbies and relevant work experience were related to high initial spatial score. Spatial ability was positively correlated with grades in both descriptive geometry and engineering drawing. Low initial ability students who later graduated made the largest gains (M = 20.55 points), those who did not graduate made smaller gains (M = 12.90 points). The authors suggested that it might be of value to teach drawing and descriptive geometry in a more concrete way for students of low (undeveloped) spatial ability whereas a more theoretical approach would be profitable for students of high ability.

Snow, Tiffin, and Seibert

In an investigation by Snow, Tiffin, and Seibert (39) learner-variable by teaching method interactions were studied. Two sections of introductory college physics classes at Purdue University totaling 437 students were divided between filmed and live physics lecture demonstrations. Nineteen short films organized into 14 classroom presentations demonstrating specific physics principles were given to one group while the other group received the parallel live teacher demonstration of the same principal. Immediate recall criteria comprised of short tests following each lecture along with delayed recall criteria comprised of questions included in five hourly course examinations were used. Using a prior knowledge of physics and 14 other learner characteristics as independent variables, a series of 2 x 3 x 3 unweighted means analyses indicated that attitude toward instructional films, ascendancy, responsibility, numerical aptitude, verbal aptitude, past experience with entertainment films, and past use of college library instructional films interacted significantly with instructional treatments (live versus filmed). This occurred primarily on the immediate recall criterion. Prior knowledge of physics modified most of these effects, however. Attitudes toward entertainment films and toward physics, emotional stability, sociability, total personality self-evaluation, academic achievement, and unspecified past experience with instructional films did not interact with treatments.

Petersen, Guilford, Hoepfner, and Merrifield

The relationship between achievement in ninth-grade algebra and general mathematics and certain factors represented in the structure of intellect model was investigated by Petersen, Guilford, Hoepfner and Merrifield (37). Approximately 600 ninth-grade students in one of the Los Angeles County high schools served as subjects. Four mathematics courses were involved, Basic Mathematics, Non-college Algebra, Regular Algebra, and Accelerated Algebra. The first two courses were listed as General Mathematics. Two achievement examinations, covering general mathematics and algebra, were given at the end of the year's work. Multiple-regression equations were computed, predicting examination achievement and final course grade from a combination of standard tests and factor tests. The following general conclusions were drawn:

(1) The higher the level of the mathematics course, the larger the number of factors that appear to be involved, from six in Basic Mathematics to ten in Accelerated Algebra.

(2) The pattern of more predictive factors varies according to the course level, but certain factors are significantly related in different courses and certain systematic differences occur as between general mathematics and algebra in their present demands on the factors.

(3) On the whole, composites of standard academic-aptitude tests and composites of factor scores do about equally well in prediction, but for somewhat different factorial reasons, with some tendency for standard tests to do better in predicting in general mathematics and factor tests to do better in predicting in algebra.

(4) When factor tests are added to each of the three standard-test combinations, multiple prediction increases significantly for algebra courses but not for general-mathematics courses.

(5) Grades and achievement-test scores are about equally predictable from either standard or factor tests, but the factor picture of course-grade criteria is usually inconsistent with that for achievement-test criteria, especially for general mathematics.

(6) Construct scoring, with detailed point systems along factorial lines, did not provide a total criterion that was any more predictable than the traditional scoring of one point for each problem correctly solved.

(7) The problem of aptitude for success in ninth-grade mathematics is even more complex than was anticipated, and will require a broader sampling of the intellectual abilities before it is solved.

(8) A test battery that will yield a high level of prediction of mathematics achievement appears to be possible, when enough of the relevant factors are included (37, p. 35).

Most of the empirical studies cited in this section of the review used variations of curricula actually being used in classrooms as instructional treatments. As a consequence the results of the studies, while supporting ATI, do not allow the isolation of relevant aspects of the curricula which are responsible for the interactions. The present research is intended to help clarify the nature of these form of content variables by analyzing existing curriculum materials for them and by using the general procedures outlined by Dunham, Guilford, and Hoepfner (14) and by Gagné and Paradise (25) for constructing treatments in which they can be varied.

III. FORM OF CONTENT VARIABLES IN TEXTUAL MATERIALS

The purpose of this series of studies was to identify form of content variables in existing textbooks that might interact with student abilities in determining achievement. It was assumed that if these variables could be identified and operationally defined, then they might be deliberately taken into account when writing textual materials. Thus, the resulting materials would have predetermined and known values on those variables which are believed to interact with learner aptitudes. Studies of the interaction hypotheses would be greatly facilitated if the investigator could control the elements in the material which are suspected to lend themselves to interactive processes.

Although there are instruments for identifying some of the abilities important in learning, no comparable ones exist for analyzing textual materials in terms of forms in which they might be cast. Therefore, these studies were concerned with the development of methodology as well as with the analysis of variables.

A. Identification of Variables; Analysis of Single Passages

Purpose

The two purposes of the study were to determine whether selected portions of textual material would be judged to depend more heavily on certain student abilities for their mastery than would other portions of the materials, and to attempt to identify form of content variables that might facilitate or inhibit learning but do not seem to be directly related to mental abilities.

Method

Subjects. Six graduate students in English education and nine graduate students in science education served as subjects.

Procedure. The procedures for both groups of students were the same but different materials were used with them. The science education students analyzed portions of the reading passages from the Glaciers Test (36). Six sections of material were selected from the reading passage. These included three narrative paragraphs, one table, one graph, and one figure.

The English education students analyzed selected passages from Chapter 27 of the textbook, Biology (35) which treats Mendelian principles of heredity. Four sections were selected from the chapter for intensive study. The first and third sections were paragraphs of written material, the second section contained three figures, and the fourth section was a table.

Both groups were instructed to read the entire passage or chapter and then to attend specifically to the designated sections. Each student received a list of brief definitions of eight cognitive abilities from the ETS Kit of Reference Tests Manual. These were Deduction, Flexibility of Closure, General Reasoning, Ideational Fluency, Induction, Number Facility, Semantic Redefinition, and Spatial Orientation. Below the statement of the definitions were blanks numbered to correspond to the selected passages in either the glaciers or biology material. Students were asked to list the mental abilities which they thought to be most importantly called into use when reading and understanding the selected passages. Also, students were asked to write in general or specific terms what they thought to be features of the materials that might facilitate or inhibit learning.

Results

The numbers of times students thought certain abilities were needed to understand the selected passages appear in Table 3.11 for both groups.

TABLE 1

Frequency with Which Abilities Were Judged Necessary for Mastering the Selected Passages

Glaciers Material

Part of Passage	Deduction	Flexibility of Closure	General Reasoning	Ideational Fluency	Induction	Number Facility	Semantic Redefinition	Spatial Ability
	Tests							
Paragraph 1	4	3	6		1			1
Paragraph 2	6	2	4		3		3	3
Paragraph 3	4	4	3		2	1	1	1
Table 1	2	4	4		2	8	2	1
Figure 1	4	1		1	1		2	7
Figure 2	4	3	6	1	5	2	1	3

Biology Textbook Material

Part of Passage	Deduction	Flexibility of Closure	General Reasoning	Ideational Fluency	Induction	Number Facility	Semantic Redefinition	Spatial Ability
	Tests							
Paragraph 1	4	5	1		3		4	1
Figure 1	4	5	2		3		3	3
Paragraph 2	4	3	5	1	1	1	1	
Table 1	2	1	4			5	1	1

A list of the facilitating or inhibiting features of the Glaciers material identified by the students appears below. It is followed by a similar list for Biology.

Glaciers

- A. Tables and graphs should be located near textual referents.
- B. Understanding something depends on prior knowledge (e.g., relation between snowfall and climate).
- C. Extremely large numbers are not meaningful.
- D. Organization is poor; lacks subheadings or other organizational devices.
- E. Passage is too "fact" oriented.
- F. Reading level (some words) may be inappropriate for ninth- or tenth-grade students.
- G. Specific examples supporting major ideas are lacking.
- H. Passage needs general overview or statement of problem or topic.
- I. Transition from paragraph to paragraph is poor.
- J. Passage evokes little interest in students.
- K. Tables are hard to understand because of use of large numbers.

Biology

- A. Use of questions
  - 1. Introductory
    - a. Can, and definitely should, be motivating.
    - b. Should lead directly into discussion.
  - 2. Interspersed, to enable student to test his understanding.
    - a. Have to be capable of being answered directly or indirectly from the material.
    - b. Need verification of answers, preferably in subtle form.

3. Summary
    a. Should be provocative but answerable.
    b. Should be at different levels (e.g., from easy to difficult, or, preferably, same questions capable of being answered on different levels by students of varying sophistication).

B. Picture-question-explanation sequence is good.

C. Leading student into doing is essential to both motivation and understanding.

D. Illustration and some explanation before attaching terms leads to less confusion than does greater density of and earlier definition of terms.

E. Informality of style is regarded as good.

F. Subheadings in boldface type are useful for transition, organization and emphasis.

G. The "What have you learned?" section is useful.

H. Format

1. Pictures are stimulating and informative.

2. Text, especially questions, should be designed to be on same or facing pages with pictures, figures, or tables.

Discussion

The frequency with which students thought that different abilities were related to mastery of the materials exhibits some clear tendencies. First, deduction, flexibility of closure, and general reasoning are most frequently cited as necessary for understanding narrative materials, and to a lesser extent are cited for figure and table presentations. Second, number facility and spatial ability are rated highly only when figure and table presentations are considered. Third, ideational fluency was mentioned rarely whether the presentations were narrations, figures, or tables.

The features which subjects pointed out as facilitating or inhibiting understanding are quite difficult to categorize meaningfully because few subjects designated the same aspects of the passages as facilitative or inhibitive. Consequently, it is inescapable to wonder whether the features which were designated were actually a result of the interaction of a particular student aptitude and an attribute of the passage. The fact that some students thought "transition from paragraph to paragraph is poor" and others thought that it was not poor actually suggests the presence of an interaction between that feature of the material and the perceiver. But in addition to features which are apparently related to cognitive aptitude, students did identify many features which they believed effected mastery but which were not cognitively related; e.g., questions should be provocative but answerable, "leading students into doing is essential to . . . motivation . . . ," "informality of style is regarded as good," and "pictures are stimulating . . ." The lack of consensus on these points and the actual occurrence of conflicting statements suggest that certain features of textual materials might interact with affective attributes of students.

B. Identification of Variables; By Comparison of Equivalent Passages

The purposes of this study were to determine whether mastery of the same content presented in three different textbooks would be judged to involve different abilities depending on the book used, and to identify other form of content variables by contrasting the ways in which the same content was presented in the three books.

In the preceding study students were asked to identify form of content variables by examining one set of materials. In the present study the same content in each of three textbooks was examined by each student. It was felt that form of content variables that existed at different levels in the three books might be more easily identified through comparisons of the three selections than by examination of only one set of materials.

Method

Subjects. Ten graduate students in educational psychology classes were used as subjects.

Materials. All students read selections on Mendelian principles of genetics from each of three high school biology textbooks. The books used and the pages read were as follows:

(a) Biological Science: Molecules to Man (3, pp. 335-345)

(b) Biology (35, pp. 411-420)

(c) Living Things (18, pp. 91-100)

Each student was given a copy of the three textbooks. Different orders of reading the selections were assigned. In addition, two mimeographed forms were given to each student. The first form contained these directions.

> Shown below is a list of 5 mental abilities. After you have read the passages, rank the books according to which would call for more of ability 1 on the part of the reader. After finishing ability 1, then rank the books on the rest of the abilities. Rank them by title names. (1 = most ability required)
>
> Definitions of Mental Abilities
>
> (a) Deduction--Ability to reason from stated premises to their necessary conclusions.
>
> _____ 1 _____ 2 _____ 3
>
> (b) General Reasoning--The ability to solve a broad range of reasoning problems including those of a mathematical nature.
>
> _____ 1 _____ 2 _____ 3
>
> (c) Induction--The ability involved in the finding of general concepts that will fit sets of data, the forming and trying out of hypotheses.
>
> _____ 1 _____ 2 _____ 3
>
> (d) Number Facility--The ability to manipulate numbers in arithmetical operations rapidly.
>
> _____ 1 _____ 2 _____ 3

(e) Flexibility of Closure--The ability to keep one or more definite configurations in mind so as to make identification in spite of perceptual distractions. Tests of this ability require the subject to search in a perceptual field containing irrelevant or distracting material in order to find one or more given configurations.

_____ 1 _____ 2 _____ 3

The second form gave the following directions.

You are asked to read the designated passages in the three books you have been given. Then on Page 1 rank the books (1 = high) according to their effectiveness in producing in students the ability to recall materials in the passages. Then identify features of the books that facilitate or inhibit this ability. Please be as specific as possible. If possible, pinpoint the sentences, paragraphs, etc., that contain the inhibiting or facilitating feature.

Examples of this ability to recall would be contained in the following test items.

(a) Define dominant-recessive trait.

(b) Pollen from tall garden pea plants was dusted on the pistils of short pea plants. One hundred peas resulting from this cross were planted the following year. All of the plants were tall. What genes for tallness did the two parent plants probably have?

_____ Tall Parent _____ Short Parent

On page 2 you are asked to rank the books and identify facilitating or inhibiting features in producing in students the ability to apply knowledge gained from the passage to new concrete situations. Again, please be as specific as possible.

An example of an item that measures this ability might be as follows:

1. In man albinism is due to a recessive gene (a); the gene for normal skin color (A) is dominant. Albinia and Smudgia are neighboring

> kingdoms. The ratio of albinos to non-albinos is about the same in both kingdoms. The inhabitants of Albinia, influenced by race propaganda, decide that the only true Albinian is one who is an albino and they set out, by drastic decrees governing matings, to produce a pure albine population. The nationalists of Smudgia, on the other hand, favor only non-albinos and enact similar decrees aimed at producing a 100% non-albino population. Which kingdom will arrive at its goal first? Explain.

Results

Each of the ranked items was analyzed by a rank sum method developed by Dunn-Rankin (15) to determine the probability of subject agreement being due to chance. The results of these analyses appear in Table 2. They

TABLE 2

Rank Sums and Results of Tests of Significance for Subject Agreement on Abilities Involved in Three Biology Textbooks*

Ability	Textbook			P
	BSCS	Biology	Living Things	
Deduction	13	21	26	BS>LT
General Reasoning	10	23.5	26.5	BS>Others
Induction	14	21	25	BS>LT
Numerical Facility	10	23.5	26.5	BS>Others
Flexibility of Closure	20	16	24	NS
Recall	13	12	17	NS
Application	8	13	15	NS

*For the five cognitive abilities a difference in rank sums between textbooks of 11 is significant at the .05 level. For recall and application only 7 and 6 subjects respectively had complete data and a difference of 9 is required for significance at the .05 level.

indicate that the BSCS material more heavily emphasizes deduction, general reasoning, induction, and numerical facility than do one or both of the other textbooks. The books apparently do not differ with regard to emphasis on flexibility of closure. Also, students did not perceive the books to differ among themselves with respect to facilitating recall or application of the material contained in them.

Facilitating and inhibiting features of the three books listed by the students are given below:

BSCS - Recall

Facilitating

A. tables, graphs and charts are good
B. headings draw attention to key ideas
C. "self check" questions helpful
D. heavy print, shaded rectangles and lines around examples are good
E. good explication of arithmetical processes
F. excellent discussion of probability and how it relates to heredity
G. progression from unknown to known
H. good for advanced students

Inhibiting

A. material too difficult for lower level students
B. too many facts
C. not enough pictures
D. too much information on probability
E. "self checks" are too limited
F. some definitions need improvement
G. material not motivating
H. too math oriented
I. lack of effective color
J. questions are vocabulary oriented

BSCS - Application

Facilitating

A. good for applying to peas
B. arithmetical processes
C. discussion of probability
D. "self checks"

E. questions at chapter end are aimed at higher levels
F. book geared to application

Inhibiting

A. too few questions
B. not interest rousing
C. "self checks" limited
D. no "why" questions
E. sticks to peas

Biology - Recall

Facilitating

A. good use of pictures
B. "useful words" and "test yourself" sections
C. good introduction
D. print type good
E. good beginning book for advanced student
F. use of italics

Inhibiting

A. illustrations seem extraneous
B. some headings are misleading
C. summary not specific enough
D. peas aren't realistic
E. pictures not directly related to text
F. definitions not entirely clear
G. poor layout

Biology - Application

Facilitating

A. good overall explanation for peas and animals
B. "problems to think about" section
C. "test yourself" and "do it yourself" sections
D. not too complicated

Inhibiting

A. some pictures not good
B. too detailed without giving pause to reflect

Living Things - Recall

A. diagrams and illustrations are explicit

B. bold type for key words
C. explanations clear
D. good use of subheadings
E. approach is direct and simple
F. easy to read
G. good for beginning student
H. illustrations easy to understand
I. book motivating
J. use of questions within the text is good
K. good color usage
L. book not "hard looking"
M. book allows for outside reading

Inhibiting

A. peas aren't realistic
B. lacks information found in other texts
C. not enough facts
D. skimpy presentation
E. rushed presentation
F. good students not challenged

Living Things - Application

Facilitating

A. "check your facts" section
B. peas to people
C. "why" questions

Inhibiting

A. questions barely adequate
B. too simple, no practice
C. very little Mendel Theory
D. presentation rushed
E. too few questions
F. no bibliography
G. inadequate coverage of material
H. coverage too superficial

C. Summary

The first activity of the project was to identify form of content variables that might interact with student aptitudes and thereby determine learning level. The period tentatively allocated to this activity was to be a major portion of the entire project. The two studies reported in this section of the report were the initial, and final, attempts at identifying these variables. Both studies revealed a great variety of features of textual materials which students believed to inhibit or facilitate learning. Some of these features, although important in their right, were of only tangential interest to the project. These features might be ascribed to poor composition of the book, such as placing the tables too far from the relevant text, or to unintended defects in the author's presentation. The other features might be categorized as having a cognitive or affective base. The latter category of features was ignored because the investigator decided at the outset to be concerned only with features that might interact with cognitive attributes of students. The remaining features, which were numerous, were of primary interest. Analyses of them revealed that the majority were highly particular and did not constitute a good basis on which to mount empirical investigations of ATI. As a consequence, the investigators chose to discontinue further studies directed to identifying form of content variables. This decision was made with some reluctance but with the expectation that a greater payoff would be achieved by following another line of attack on the problem. Essentially, this task was to attempt to identify empirically ATI effects in learning by students, classified according to aptitude, through exposure to parallel sets of materials which contain different known values of a particular form of content variable, embody clearly identified teaching strategies, or heavily reflect one or more aptitudes. The subsequently reported studies of redundancy of instructional materials, vocabulary learning, and mathematical operations are examples, respectively, of the learning materials.

IV. REDUNDANCY AS A FORM OF CONTENT VARIABLE

Introduction

Readability or reading difficulty of textual material is one form of content variable which has been widely investigated. But it has not been studied in the context of the aptitude treatment interaction hypothesis. However, it seems evident that persons having low reading skill and verbal ability will be more successful in treatments that involve materials of low reading difficulty rather than high reading difficulty; although it is not as evident that persons having high verbal ability and reading skill will achieve best on treatments involving high reading difficulty. In addition to the expected interaction of verbal ability with difficulty level of reading materials, it is possible that abilities other than verbal ability are involved in differential achievement under two levels of reading difficulty.

The four studies reported here deal with redundancy as measured by the cloze procedure. The purpose of the first study was to identify variables in textual material that determine its redundancy level. Knowledge of these variables might enable the construction of sets of material with known redundancy levels. The second study is an investigation of relationships between a set of cognitive abilities and redundancy of materials from several contents. The third and fourth studies were intended to determine whether passages from graded-readers vary in redundancy and, if they do, whether evidence for the aptitude treatment interaction hypothesis might be found through use of them.

The term "redundancy" is used here in a general sense to indicate the amount of sequential constraint or organization among the words in a passage and to denote the extent to which a given word of a passage is determined by context. If a word is deleted but is so completely determined that it can be restored, it is considered to be redundant. The relative redundancy of the passages was estimated by the cloze procedure. Cloze scores for passages are derived by averaging the proportion of words which subjects can correctly restore to mutilated passages; e.g., one in which every fifth word has been deleted.

The cloze procedure, as do many other methods of assessing redundancy, provides an indirect estimate of the relative word redundancy of passages from the word restorations of individuals. It was selected because more information about its reliability and validity is available than for other techniques.

It was postulated that achievement from exposure to redundant and nonredundant instructional materials would be mediated by cognitive aptitudes. The studies which are presented here are necessary preliminaries to investigating the hypothesized influence of cognitive aptitudes.

A. Redundancy, Content Area and Expressional Style*

The central problem treated in this study arose when the absence of guidelines for writing instructional passages of known redundancy values was noted. The redundancy of passages is known to be related to their readability. In turn, the assessment of readability depends on style factors in the passage. Because one might write instructional materials in different expressional styles while simultaneously preserving the original content, manipulating style appeared to be a logical means of varying redundancy if relationships between individual elements of expressional style and redundancy could be determined. Thus, the specific task was to determine relationships between redundancy and individual elements of expressional style.

The elements of style which were studied had been identified previously by other researchers. These were:

---

*Beard, J.G. The Relation of Redundancy to Content Area and Expressional Style in High School Textbooks. Unpublished doctoral dissertation, Florida State University, 1966.

monosyllable words, sentence length, first person pronouns, second person pronouns, third person pronouns, different words, complex sentences, infinitive phrases, words beginning with "1", and a measure of abstraction. Relationships between redundancy and the extent of use of each of these expressional elements were determined for sample instructional passages.

Previous research indicated that the expressional style of materials might differ between subject matter content areas. Therefore the relationships between redundancy and style were determined separately for each of four content areas of biology, chemistry, American government, and world history as well as for the combined group of the four content areas.

Method

Subjects. The subjects were 250 beginning tenth-grade students from a public school. This grade-level was chosen because students would have had relevant prior instruction in the four content areas but would not have studied any of the areas in depth.

To ensure that the sample was representative of tenth-grade students, 50 students were selected from each of the Florida-based percentile ranges 1-19, 20-39, 40-59, 60-79, and 80-99 on the verbal section of the School and College Ability Test (SCAT) which had been administered to the students one year earlier as part of the Florida State-wide Ninth-grade Testing Program. Each cloze test form was administered to equal numbers of students in each ability level.

Materials. Textbooks were sampled from all comprehensive high school textbooks in print in the content areas of general biology, chemistry, world history, and American government. Programmed textbooks, pictorial books, workbooks, or other books in nonprose formats were eliminated from the list before the sample was chosen. A "comprehensive" textbook is one which deals with the whole content, and it stands in contrast to specialized books which are used as collateral texts. The procedures which were used to select the contents, textbooks, and passages, and to construct the cloze test forms are described below.

The four contents--biology, chemistry, world history, and American government--were arbitrarily chosen. Each meets the following criteria which were established as requisites for inclusion in the study: (a) The content must be generally taught in high schools throughout the United States; (b) The content must be so limited in scope that a single book is usually used as a main text in the high school level; (c) The content must be suitable for both boys and girls; and (d) Ten or more textbook titles must appear in Textbooks in Print 1965 (41) for that content. (This is necessary to provide the sample size required by the experimental design.)

Ten textbooks were randomly selected for each of the four contents. Textbooks in Print 1965 (41) was examined and titles falling in the populations defined above were numbered consecutively for each content separately. Ten of the books were then selected randomly without replacement for each content.

An instructional passage was chosen from the main body of each textbook according to the following procedure. A table of random numbers was used to select a page number in each book. A search was begun at the top of the page selected and carried forward until a subtitle was located. A subtitle was regarded as any section or subsection heading which is lower in rank than a chapter heading. The passage following the subtitle was used if it met the following criteria: (a) The passage did not duplicate a topic previously chosen for the study; (b) The passage contained at least 250 words of continuous prose (A passage was considered discontinuous if the reader was directed to material outside the passage; e.g., if he was directed to figure 1.); and (c) The 250 words did not include more than 10 quoted words. When the topic chosen did not meet the criteria, another topic was chosen by the same procedure.

The first 250 words of continuous prose following the subtitle constituted the sample passages.

The 40 sample passages were analyzed by a special analyst to determine the extent of use of the following elements of expression:

A. Nouns of Abstraction--The abstractness of the sample passages was estimated by counting

Gillie's "nouns of abstraction." These are words having the suffixes; ness, ment, ship, dom, nce, ion, and y.

B. Monosyllable words--The number of monosyllabic words in each passage.

C. "1" words--The number of words beginning with the letter "1" in each passage.

D. First person pronouns--The number of first person pronouns in each passage.

E. Second person pronouns--The number of second person pronouns in each passage.

F. Third person pronouns--The number of third person pronouns in each passage.

G. Different words--The total number of words in the passage minus the number of repeated words.

H. Sentence length--The mean number of words per complete sentence in each passage.

I. Complex sentences--The proportion of complete sentences having one main clause and one or more subordinate clauses.

J. Infinitive phrases--The total number of infinitive phrases in each passage.

The cloze test forms were constructed by a systematic deletion process. Every fifth word in the sample passages was replaced by an underlined blank space, one inch in length. The deletion process began with the first word of the first sentence. Four other forms were prepared for each passage by beginning the deletion process on the second, third, fourth, and fifth words respectively. By using these five forms for each passage, every word was included as an item to be restored. This procedure provided 200 different test forms (four contents represented by ten books each, and each book represented by five parallel cloze tests).

The cloze test forms were typed on offset mats and reproduced on 8 1/2 by 14 inch paper, one test form per sheet. The title was centered at the top of each sample passage and was not mutilated. Four test forms, one for each content, were bound into a booklet having an attached cover sheet of instructions. These instructions and the 40 forms appear in Appendix A. Within each booklet, the forms were arranged randomly to control for practice effects.

Administration of the Cloze Test. The tests were administered to students in intact classroom groups. The students were advised to follow the administrator's reading of the instructions. Student questions were then answered. They were advised that a bell would sound at ten minute intervals. Each time the bell sounded, the students were told to begin work on the next test form. The ten minute periods were used because on a pilot administration 16 out of 20 high school students completed a similar form in eight minutes. A check of the completed test forms of the present study revealed that 82 percent had a response in at least one of the last three blanks.

Scoring the Test Forms. To be scored as correct, responses had to match the words deleted except for minor misspellings; i.e., synonyms were not considered to be correct. The cloze score for each passage was derived by averaging the percentage of words correctly restored on its cloze test forms. This score became a basic measure for the analyses.

Analysis. For each content separately, means and standard deviations were computed for the cloze and expressional element scores. The significance of differences among the mean scores over passages within contents was determined through a one-way analysis of variance procedure. Homogeneity of the variances of the cloze and expressional element scores within each content was determined using Hartley's F-Max test for homogeneity of variance, according to procedures outlined by Winer (45).

Relationships between passage cloze scores and individual elements of expressional style were determined through correlational analyses.

The scores for all passages in the four contents were then combined into a single distribution. The correlations between cloze scores and each element of expression were recomputed using the entire distribution of 40 passages.

A multiple regression analysis was performed, using the cloze scores as the criterion variable and the expressional elements as the independent variables. The multiple correlation coefficients and squared multiple correlation coefficients were computed to show the amount of overall relationship between the cloze scores and the expressional elements.

The individual restoration scores were analyzed using a four variable factorial design. The model for the analysis was the following:

$$Y_{ijkl} = M+C_i+P_{j(i)}+F_{k(ij)}+V_l+CV_{il}+PV_{j(i)l}+FV_{k(ji)l}+G_{ijkl}$$

where Y_{ijkl} = an individual restoration score and

M = grand mean of restoration scores

C_i = effect of contents 1, 2, 3, 4

$P_{j(i)}$ = effect of passages within Contents 1, 2, ... 10

$F_{k(ij)}$ = effect of deletion forms within Passages and Contents 1, 2, ... 5

V_l = effect of ability level of subjects 1, 2, ... 5

and $n = 1$

The null hypotheses were: (a) Content effects are zero, (b) Passage effects are zero, (c) Ability level effects are zero, and (d) The interaction effects are zero.

The ratios involved in analysis of variance for the several hypotheses were: (a) mean square for content means to the passages (within contents) mean square, (b) mean square for passage means to the forms (within passages and contents) mean square, (c) mean square for ability means to the passage by ability level interaction mean square, (d) mean square for content by ability level interaction to the passage (within contents) by ability level interaction mean square, and (e) mean square for passage (within contents by ability level interaction to the form by ability level interaction).

Results

Table 3 contains the means and standard deviations of the passage cloze scores. The significance of differences among means for the content areas were examined using one-way analyses of variance. Also, the homogeneity of variances were tested by Hartley's F-Max test. The F-ratios and F-Max ratios appear in Table 3.

Product moment correlation coefficients were computed between the cloze scores and the elements of expression for each content over passages within the content. The correlation coefficients appear in Table 4.

Significant relationships appear between the cloze scores and two or more of the expressional elements for each content area.

The relationships between the cloze scores and four of the expressional elements are consistent in direction across the four content areas. The elements are monosyllables, first person pronouns, different words, and abstract nouns. Only complex sentences bear no appreciable relationship with the cloze scores in any of the content areas.

The ten sample passages in each content area were combined into a single distribution of forty passages. Product moment correlation coefficients were then computed among the cloze scores and elements of expression in the combined distribution. The intercorrelation matrix appears in Table 5.

The correlations between the cloze scores and each of monosyllables, sentence length, first person pronouns, different words, and nouns of abstraction were significant at the .10 level. Monosyllables and second person pronouns were positively related to cloze scores; sentence length, first person pronouns, different words, and nouns of abstraction were inversely related to the cloze scores.

The vocabulary difficulty factor, as measured by monosyllabic words, is the best single predictor of redundancy. This was expected because vocabulary

TABLE 3

Means, Standard Deviations, and F-Ratios For Passage Cloze Scores and Expressional Elements

Passage Variable		World History	American Govt.	Biology	Chem- istry	F-Ratio
Cloze Score (%)	$\bar{X}$	35.4	35.3	35.6	38.4	.5
	S	5.3	8.3	8.2	3.3	6.5
Number Monosyllable	$\bar{X}$	152.2	150.0	157.7	148.7	.6
	S	21.6	16.8	16.0	10.1	4.5
Sentence Length	$\bar{X}$	17.6	17.8	17.1	16.6	.3
	S	1.9	3.5	3.0	3.2	3.6
Different Words	$\bar{X}$	151.4	152.0	144.3	147.2	.7
	S	8.6	10.9	17.1	16.6	3.9
Number "i" Words	$\bar{X}$	11.2	13.1	18.8	17.6	3.9
	S	4.0	3.6	7.2	7.4	4.2
Nouns of Abstraction	$\bar{X}$	5.7	6.7	6.1	3.6	1.4
	S	4.9	3.4	3.1	3.0	3.7
First Person Pronouns	$\bar{X}$	.6	.9	1.3	.5	.8
	S	1.1	1.9	1.2	.7	7.3
Second Person Pronouns	$\bar{X}$	.2	1.5	1.3	.4	.7
	S	.4	4.4	2.2	.7	39.5
Third Person Pronouns	$\bar{X}$	9.9	7.0	6.0	5.1	2.0
	S	6.3	5.3	4.0	1.9	11.6
Complex Sentences (%)	$\bar{X}$	29.5	30.2	43.1	37.9	2.2
	S	13.2	10.6	16.9	14.2	2.5
Infinitive Phrases	$\bar{X}$	6.3	4.7	3.0	2.9	4.7
	S	2.9	2.6	1.6	2.1	3.1

TABLE 4

Correlations Between the Elements of Expression and the Cloze Scores for the Passages Within Each Content Area

N = 10 passages

Elements of Expression	World History	American Govt.	Biology	Chemistry
Monosyllables	.66*	.68*	.61*	.31
Sentence Length	.40	-.68*	.00	-.75*
First Person Pronouns	-.10	-.10	-.84	-.35
Second Person Pronouns	.17	.65*	.14	-.44
Third Person Pronouns	.13	.29	.60*	-.32
Different Words	-.22	-.44	-.17	-.41
Infinitive Phrases	.64*	-.24	.34	-.36
"i" Words	-.64*	.02	.28	.60*
Abstract Nouns	-.25	-.41	-.37	-.33
Complex Sentences	-.01	-.01	.14	-.16

*significant at .10 level

TABLE 5

Correlations Between Cloze Scores and Expressional Elements for Forty Sample Instructional Passages

Expressional Element	2	3	4	5	6	7	8	9	10	11
Cloze Scores	.54*	-.32*	-.33*	.35*	.21	-.27*	.03	.15	-.36*	-.01
Monosyllables		-.17	-.22	.31	.31	-.24	.17	.17	.07	.06
Sentence Length			.10	-.16	-.03	.32	.33	-.30	.17	.40
First Person Pronouns				.04	.00	-.01	.16	-.08	-.11	.10
Second Person Pronouns					-.13	-.09	.08	.09	-.04	.06
Third Person Pronouns						.17	.21	-.28	-.19	.16
Different Words							.12	-.39	.08	.16
Infinitive Phrases								-.40	.02	.40
"i" Words									-.20	-.17
Nouns of Abstraction										-.07
Complex Sentences										

*significant at .05 level

difficulty has repeatedly been shown to be the most important factor in studies of readability. The next best predictor was "nouns of abstraction" followed closely by the indices of human interest, first and second person pronouns, and the sentence difficulty factor, sentence length. Different words, a second measure of the vocabulary factor, shows the least significant relationship with the cloze scores.

Multiple correlations between the cloze scores and expressional elements appear in Table 6.

TABLE 6

Summary Table
Multiple Regression Analysis

	Variable Entered	Multiple R	R^2	Increase in R^2	F Value	P
1.	Monosyllables	.5439	.2958	.2958	15.9624	.05
2.	Abstract Nouns	.6067	.3681	.0723	4.2352	.05
3.	First Person Pronouns	.6620	.4382	.0701	4.4930	.05
4.	Second Person Pronouns	.6997	.4896	.0513	3.5200	.10
5.	Sentence Length	.7147	.5107	.0212	1.4709	
6.	Complex Sentences	.7237	.5238	.0130	0.9035	
7.	Different Words	.7299	.5327	.0090	0.6139	
8.	Third Person Pronouns	.7370	.5432	.0104	0.7080	
9.	Phrases	.7392	.5465	.0033	0.2177	
10.	"i" Words	.7402	.5479	.0014	0.0902	

N = 40 passages

The coefficients were obtained by systematically combining the expressional elements in the order of their relative contribution to the relationship. A multiple R of .70 was obtained between the cloze scores and monosyllables, abstract nouns, first person pronouns, and second person pronouns. The multiple R^2 indicates that these four elements account for 49 percent of the variance in the cloze scores. None of the remaining elements added a significant increment to the multiple R at the .10 level of confidence.

The results of the analysis of individual restoration scores appear in Table 7. The error term used in making each of the F tests was described in the method of analysis section.

TABLE 7

Analysis of Variance of Individual Restoration Scores

Source	df	MS	F	P
Contents (C)	3	138.71	.51	NS
Passages within Contents (P w. C)	36	273.27	6.15	.01
Forms within Passages and Contents (F w. C & P)	160	44.44		
Verbal Ability Level (V)	4	5,405.71	159.84	.01
CV	12	6.37	.19	NS
V x (P w. C)	144	33.82	1.95	.01
V x (F w. P & C)	640	17.34		
Total	999			

The restoration scores did not discriminate significantly among the content areas. Significant differences were found among the means for passages and for verbal ability levels. In addition, a significant passages (within contents) by verbal ability level interaction effect was found.

The mean restoration scores for the verbal ability levels appear in Table 8. The means reveal a marked increase at each successive level of verbal ability, with slightly greater increases at the lower end of the verbal ability scale.

TABLE 8

Means and Standard Deviations of Restoration Scores for Subjects at Five Ability Levels

SCAT Percentile	Mean Restoration	Standard Deviation	N
01-19	43.62	14.53	50
20-39	61.62	13.28	50
40-59	74.20	12.86	50
60-79	85.70	16.06	50
80-99	96.96	16.22	50

The significant passage by ability level interaction indicated that the mean restoration scores for the passage by ability level combinations could not be predicted from the passage and the ability level main effects alone. To facilitate the observation of interaction effects associated with passage difficulty, the 40 passages were divided into two groups; the first group consisted of the 20 passages having cloze scores above the grand mean of the cloze scores, and the second group consisted of the 20 passages at and below the grand mean. Mean restoration percentages were computed for each of the five ability levels over the passages in each of the two groups.

These mean restoration percentages were then adjusted to remove the variation attributable to passage difficulty and the variation attributable to verbal ability. The adjusted mean restoration percentages appear in Table 9.

TABLE 9

Mean Restoration Percentages with Variations Associated with Passage Difficulty and Verbal Ability Level Removed

Passage Difficulty Level	Verbal Ability Level 1	2	3	4	5	Grand Mean
Easy Cloze	35.5	35.3	36.7	36.7	36.4	36.2
Difficult Cloze	36.9	37.0	35.6	35.9	36.0	

The adjusted mean scores for ability levels 1 and 2 tend to be lower in the easy passages and higher in the difficult passages than the value predicted from the main effects alone, 36.2. The adjusted means for ability levels 3, 4, and 5 tend to be higher than expected in the easy passages and lower in the difficult ones.

The results show that six elements of expressional style are related to the redundancy of high school instructional materials. These expressional elements are monosyllable words, sentence length, nouns of abstraction, different words, first person pronouns, and second person pronouns.

The findings reveal that instructional materials are more redundant if they contain relatively *large* numbers of monosyllable words, *large* numbers of second person pronouns, *small* numbers of "nouns of abstraction," *small* numbers of different words, *small* numbers of first person pronouns, and a relatively *short* mean sentence length.

The overall relationship between the measures of redundancy and the extent of use of these six expressional elements is indicated by a multiple correlation coefficient of .72.

Discussion

Relationships were found between redundancy and expressional elements over the passages within each content area; however, the predictive value of each expressional element varied considerably among the contents. Consistencies in direction of relationship across the four contents were found for monosyllables, first person pronouns, different words, and abstract nouns. However, these relationships varied widely in amount. The remaining six elements varied in direction of relationship across the contents as well as in amount. No rationale could be advanced to account for the different patterns of relationships among the content areas.

The restoration scores did not discriminate among the four content areas. This finding indicates that the redundancy of instructional materials is not associated with the broad subject matter areas.

The restoration scores were highly and positively associated with verbal ability. Also, the scores of the high ability subjects provided a finer discrimination between the passages of high and of low redundancy.

It was not possible to test for the effect of forms within passages and contents. The amount of error caused by using only a single form would decrease as the number of cloze blanks became larger and as the number of subjects became larger; but it is not known at what point the errors would become insignificant.

The restoration scores did discriminate among the sample instructional passages. However, the results of this study do not enable drawing conclusions about whether differences in redundancy among passages is related to the particular topic selected or to style of the authors because forty different topics by forty different authors were used. It is conceivable that the redundancy of each passage reflects the style of its author.

B. The Relation of Cloze Scores to Selected Cognitive Variables*

The major purpose of this study was to investigate the relationships between cloze scores for individuals and a set of relatively factor-pure cognitive tests. Classification of the abilities involved in the cloze procedure would appear to be useful in increasing the theoretical and practical potential of the cloze procedure when used in aptitude treatment interaction studies. A minor purpose of the study was to investigate the interrelationships of the cloze scores.

---

*Kohler, E. T. An Investigation of Cloze Scores

Method

Subjects. The subjects were 257 tenth-grade students, 128 males and 129 females, from one senior high school. Complete sets of cloze test and "factor-pure" ability data were obtained from 243 of the students. There was no attempt to restrict the range of ability and achievement of this sample because it was desired to have a group of tenth-grade subjects with as much heterogeneity as is found in a typical tenth-grade.

Materials. The passages used in the construction of the cloze tests were selected from the science and social studies passages mentioned previously. There were two biology (B5 and B8) and two chemistry (C6 and C7) passages used in the science area. For the social studies tests, two American Government (A3 and A6) and two World History (W1 and W8) passages were used.

Because of the literature on the cloze procedure does not provide a useful guide for selecting a battery of "factor-pure" cognitive ability tests, the selection of these tests was to a considerable extent a function of logical judgments about what abilities were demanded by the materials.

A pilot study was done with 30 eleventh-grade students. Results suggested that the cloze task is highly verbal but that other special abilities might play an important part. Based on this study, the Map Planning Test which gives a measure of spatial scanning, and the Cube Comparisons Test which gives a measure of spatial orientation (22), were added to the battery.

A predictor battery having low test intercorrelations was desired for use in the multiple

---

in Terms of Selected Cognitive Variables. Unpublished doctoral dissertation. Florida State University, 1966.

regression analysis. It was expected that by using factor analysis an almost orthogonal battery of "factor-pure" tests could be devised.

The students used in this factor analytic selection procedure were 250 tenth-grade students who were part of the sample in a study conducted at the Florida State University (36). In that study, a battery of thirty-three tests which measured fourteen reference factors (22) were administered to approximately 1,200 public school students in the ninth- through twelfth-grades. The thirty-three variables were intercorrelated and a series of principle axis solutions and varimax rotations were executed. The matrix resulting from the rotation of fourteen factors appeared to be the most interpretable. This matrix, the listing of the reference factors, and the tests used to measure them are reported elsewhere (34).

Of the fourteen factors which were rotated, nine were interpreted. These nine factors served as the basis for selecting additional"factor-pure"tests. Each interpretable factor was examined in turn and the test having the highest loading and most purity for a given factor was selected (Table 10).

Each student received five scores for each cloze test. The first three of these were most important: (a) A total score. The total number of deletions correctly restored. The results and discussion of the results will be in relation to this score unless otherwise specified. (b) A content word score. The number of content words correctly replaced. Content words were defined as nouns, verbs, adjectives, adverbs and gerunds. (c) A connective word score. The number of connective words correctly replaced. Connective words were defined as conjunctions, prepositions, auxiliaries, and articles.

In addition to these three scores, an "odd" words score and "even" words score were obtained. To obtain these scores the deletions were numbered from 1 to 50. The "odd" score was the number of odd numbered deletions replaced, and the "even" score was the number of even numbered deletions replaced. These two scores were used only to calculate reliabilities of the cloze tests.

TABLE 10

Results of Factor Analysis Test Selection Procedure

Test	Factor Represented	Factor Name
Inference Test	I	Reasoning
Addition Test	II	Number Facility
Object Synthesis	III	Ideational Fluency
Hidden Patterns Test	IV	Flexibility of Closure
Planning Air Maneuvers	V	Figural Adaptive Flexibility
Picture-Number Test	VII	Associative Memory
Wide Range Vocabulary Test	VII	Verbal Comprehension
Figure Classification	X	Induction
Logical Reasoning	XII	Reasoning

Results

The analyses are divided into two categories; the description of the variables, and the description of the relationships between the variables or sets of variables. The first part of this section presents the results which describe the characteristics of the cloze tests and the battery of "factor-pure" tests. Then the results that describe the relationships between the "factor-pure" tests and the cloze tests are presented.

The means and standard deviations for the three types of scores obtained from each of the eight cloze tests and the mean percentage of correct responses appear in Table 11.

TABLE 11

Cloze Test Means, Standard Deviations, and Mean Percentage Correct

Test	Total Score			Content Word Score			Connective Word Score		
	$\overline{X}$%c	$\overline{X}$	SD	$\overline{X}$%c	$\overline{X}$	SD	$\overline{X}$%c	$\overline{X}$	SD
B5	49	23.91	5.82	46	11.61	3.54	50	12.45	2.88
B8	31	15.63	6.10	10	3.13	2.29	62	12.51	4.41
C6	38	19.04	5.57	33	11.58	3.65	50	7.53	2.61
C7	24	12.19	5.52	21	7.25	3.10	31	4.96	2.92
A3	25	12.72	5.92	14	3.98	2.65	40	8.95	3.84
A6	39	19.40	6.34	37	8.19	3.58	58	11.60	3.11
W1	28	13.85	4.31	11	3.42	2.08	52	10.44	2.84
W8	36	17.87	6.89	27	9.61	5.14	59	8.25	2.36

Table 12 contains the reliabilities of the eight cloze tests.

Two factor analyses were computed on the cloze scores. In each analysis unity was placed in the diagonal cells as estimates of the communalities and all factors with eigenvalues greater than one were rotated.

TABLE 12

Cloze Test Reliabilities

Test	r
B5	.91
B8	.87
C6	.81
C7	.84
A3	.86
A6	.88
W1	.79
W8	.89

The first analysis was of the eight cloze total scores. Examination of the factor matrix revealed one large factor (eigenvalue of 4.90) which accounted for approximately 61 percent of the total variance.

The second factor analysis pertained to the eight content word scores and the eight connective word scores. It revealed three factors with eigenvalues greater than one. In this analysis, as in the first, one large factor emerged which accounted for approximately fifty percent of the total variance. The eigenvalue associated with this factor was 8.04. The other eigenvalues greater than one were 1.09 and 1.01. The second and third factors accounted for seven percent and six percent of the variance respectively. These three factors were rotated and the rotated factor matrix appears in Table 13.

There were no analyses specific to the "factor-pure" tests but as a part of other analyses the intercorrelations between these predictor tests were found. These are presented in Table 14.

To explore the relationships between cloze and "factor-pure" tests stepwise multiple regression coefficients were calculated. Each of the cloze scores for science, social studies, and a total were used as dependent variables, and the eleven "factor-pure" test scores were used as independent variables for the

TABLE 13

Factor Analysis of Cloze Content and Connective Scores; First Three Rotated Factors

Test		I*	II	III
B8	Content	408	434	460
B8	Connective	290	517	547
B5	Content	548	242	447
B5	Connective	500	162	512
C6	Content	253	180	718
C6	Connective	172	075	828
C7	Content	724	295	275
C7	Connective	607	404	313
A3	Content	784	133	179
A3	Connective	731	150	313
A6	Content	381	753	086
A6	Connective	081	826	176
W1	Content	375	644	213
W1	Connective	312	419	506
W5	Content	724	373	209
W5	Connective	694	265	189

2

TABLE 14

Intercorrelations of Factor-Pure Tests

Test	2	3	4	5	6	7	8	9	10	11
1. Addition Test	500	184	204	152	125	009	298	133	105	082
2. Map Planning Test		146	169	182	178	164	319	212	050	027
3. Wide Range Vocabulary Test			002	237	137	211	200	054	133	292
4. Picture-Number Test				176	113	038	138	146	-003	009
5. Object Synthesis					161	200	271	162	089	235
6. Planning Air Maneuvers						129	129	085	061	141
7. Logical Reasoning							237	230	271	359
8. Hidden Patterns Test								312	290	181
9. Cube Comparisons Test									278	273
10. Figure Classification										186
11. Inference Test										

*decimals have been omitted

computation of three regression equations. The results of these analyses appear in Table 15. The relative contribution of each "factor-pure" test for predicting the cloze scores is shown in Table 16.

Discussion

The results indicate that Wide Range Vocabulary Test, Logical Reasoning, and Inference Test were consistently and significantly related to the cloze tests. The next most frequently significant tests were Addition Test and Hidden Patterns Test. Planning Air Maneuvers, Map Planning Test, Object Synthesis, and Picture-Number Test, Cube Comparisons Test and Figure Classification appeared to be relatively unrelated to the cloze tests used.

A secondary purpose of the study was to examine the interrelationships of the various cloze scores. This was done by means of two factor analyses. These analyses indicate the presence of a general factor in the set of eight cloze total scores and in the set of sixteen content and connective word scores.

From the results of the present study several conclusions were drawn. The cloze task was relatively independent of the content covered by the tests. Little additional information was gained from the use of content and connective word scoring schemes. There was a general factor present in the cloze scores. The analyses of the cloze tests in terms of the "factor-pure" tests provided information about the abilities which are required to complete successfully the cloze task. The tests which were used to measure these abilities are presented below in the order of their importance for predicting cloze scores.

(a) Wide Range Vocabulary Test. This was thought to be a reflection of the subject's knowledge of grammatical rules and breadth of vocabulary. It was considered as the foundation ability in the cloze task.

(b) Logical Reasoning. This was believed to be a somewhat formal reasoning ability which was thought to be related to the use of rather rigid sets of rules to solve problems.

TABLE 15

Summary of Step-Wise Regression Analyses

Independent Variables	Dependent Variables Science Total Raw Weights	 Social Studies Total Raw Weights	 Total Raw Weights
Wide Range Vocabulary Test	1.014*	1.034*	2.048*
Logical Reasoning	1.052*	.847*	1.899*
Inference Test	.620*	.716*	1.336*
Hidden Patterns Test	.431*	.171	.601*
Planning Air Maneuvers	.283	.238	.521*
Map Planning Test	.219*	.156	.375*
Cube Comparisons Test	.091	.146	.237
Addition Test	.192	.274*	.466
Object Synthesis	.100	.098	.198
Picture-Number Test	-.181	.058	-.123
Figure Classification	-.070	-.133	-.203
R	.715	.664	.717
Intercept	12.440	19.740	32.180

*Variable made a significant contribution to the increase in R^2 at the .05 level or greater.

TABLE 16

Factor-Pure Tests Ranked as to Proportion of Variance Accounted for in Criterion Variable

Independent Variable	Biology	Chemistry	American Govt.	World History	Science	Social Studies	Total Cloze
Logical Reasoning	1*	2*	2*	2*	1*	2*	2*
Wide Range Vocabulary Test	2*	1*	1*	1*	2*	1*	1*
Inference Test	3*	4*	3*	3*	3*	3*	3*
Hidden Patterns Test	4*	3*	11	5	4*	6	5*
Addition Test	5*	11	4*	4*	7	4*	7
Planning Air Maneuvers	6	7	5	6	6	5	6
Map Planning Test	7	5	7	7	5*	7	4*
Picture-Number Test	8	8	8	11	8	11	11
Cube Comparisons Test	9	10	6	10	10	8	9
Object Synthesis	10	6	10	9	9	10	8
Figure Classification	11	9	9	8	11	9	10

*Variable made a significant increase in R^2 at .05 level or greater.

(c) Inference Test. This measure of reasoning was thought to reflect a less formal, more flexible approach to the problem of the cloze task.

(d) Addition Test. The ability measured by this test was thought to be important because this speed factor provided the subject with more exposure to the contextual clues found within the cloze test.

(e) Hidden Patterns Test. This measure of the ability to cast aside given solutions in favor of better ones was considered to be important because there can be many correct words used to replace the deletions but to be successful the subject must at times cast aside these technically correct solutions in favor of ones that are both technically correct and correct in relation to the contextual restraints imposed by the passage.

(f) In relation to the above "factor-pure" tests the following appeared to be less important abilities in the cloze task: Figural Adaptive Flexibility, Spatial Scanning, Semantic Redefinition, Associative Rote Memory, Spatial Orientation, and Induction.

C. Relative Redundancy of Reading Material with Different Readability Levels and Parallel Content

The purpose of this study was to derive experimental instructional passages for use in testing an aptitude treatment interaction hypothesis. The particular form of content variable was redundancy as measured by the cloze procedure. Previous studies revealed that cloze scores of instructional materials were greater when the vocabulary was simple, the sentences were short, and few abstract words were used.

An examination of existing reading materials led to the identification of a graded reading series having parallel versions at different levels of linguistic complexity (5,6). The two versions of each passage cover the same content but one version has greater vocabulary range, greater sentence complexity, and longer sentences and paragraphs. These two versions are referred to subsequently as the "original" and the "classmate" editions. Four reading passages were selected and the relative redundancy levels of the two versions of each were determined. The stories selected from the original edition are reported by the publisher to be at approximately the sixth-grade reading level. The parallel versions from the classmate edition are reported by the publisher to contain 40% fewer new words, 33% fewer running words, shorter sentences, and shorter paragraphs than the original versions. It was hypothesized that the passages from the classmate edition would be more redundant than those from the original edition.

Method

Subjects. The subjects were 81 male and female sixth-grade pupils from a predominately Negro public school. Ability data had been collected from them six months prior to this study.

Materials. Eight cloze tests were prepared from four stories contained in the elementary reader entitled Stories to Remember and the corresponding versions of these four stories contained in the classmate edition of the same reader. The stories were: "The Shining Metal," "Adventures of the Whalers," "Independence Day," and "The Pony Express Rider." The eight cloze tests consisted of the first 300 words of each story. Every fifth word of the text in each cloze test passage was omitted and was replaced by a blank line 15 typewriter spaces long. Each cloze test contained blanks for 60 words.

The cloze tests were organized into four types of test booklets. Each booklet contained four different passages, representing each of the stories; two were from the classmate edition and two were from the original edition. The stories were arranged in booklets according to the version as shown in Table 17. A complete set of the materials is shown in Appendix A.

Administration of Cloze Tests. The test booklets were distributed randomly to the students.

TABLE 17

Story Arrangement in Cloze Test Booklets

Story	Booklet Type I	II	III	IV
The Shining Metal	Original	Classmate	Original	Classmate
Adventures of the Whalers	Classmate	Original	Original	Classmate
Independence Day	Original	Classmate	Classmate	Original
The Pony Express Rider	Classmate	Original	Classmate	Original

Instructions for completing the cloze tests were read aloud to the students. They were told to: (a) read the complete passage at least once before filling in any of the blanks, (b) try to fill in the blanks with the words you think have been omitted, (c) try to complete as many words as possible even if you have to guess, (d) reread the completed passage and change any of your word entries if you can think of a better word to replace it.

The students completed the first two cloze tests on the first day and completed the remaining two tests two days later. All students completed the passages in the same sequence: "The Shining Metal," "Adventures of the Whalers," "Independence Day," and "The Pony Express Rider." Twenty-five minutes were allowed to complete each passage.

The cloze tests were scored by comparing the subjects' responses with the exact words used in the published text. Only legible words which exactly matched the words used in the published text, or minor misspellings of the exact words, were scored as correct. The mean number of words correctly restored to each passage was used as the passage cloze score.

Analysis. The means and standard deviations of the cloze test scores for each version of the four stories were computed and the difference between means on the classmate version and the original version of each story was tested for the significance.

Results

The means and standard deviations of the cloze scores for each passage are shown in Table 18. The mean scores on the classmate versions were significantly higher ($p<.01$) than those on the original version for three of the four stories. "Adventures of the Whalers" was the exception; on that story the mean cloze scores were higher for the original version. Analysis of the two versions of the whaling story indicated no apparent reason for this reversal related to their expressional style, rather it appeared that the low classmate cloze score resulted from the vagaries of sampling in that the words deleted from the classmate version contained an inordinately large number of key content words which are more difficult to restore than connective words.

TABLE 18

Means and Standard Deviations of Cloze Scores

	The Shining Metal		Adventures of the Whalers		Independence Day		The Pony Express Rider	
	Original	Classmate	Original	Classmate	Original	Classmate	Original	Classmate
Mean	12.44	22.35	18.23	11.95	14.60	20.66	16.55	21.51
SD	7.46	9.23	4.80	6.64	6.94	9.81	7.53	8.44
N	41	40	40	41	40	41	42	39

Discussion

The data indicate that the alternate versions of three of the four stories selected were written at different levels of redundancy with the classmate version being more redundant. No attempt has been made to express the amount of redundancy in other than relative terms.

The content of the two versions is similar for each of the four stories studied; however, a thorough content analysis would require the evaluation of subtle differences in expression. Such an analysis has not been made.

In summary the three stories--"The Shining Metal," "Independence Day," and "The Pony Express Rider"--were regarded as suitable passages for use in investigating the interaction of learner aptitude and redundancy level of instructional material.

D. Reading Comprehension as a Function of the Interaction of Learner Aptitude and Redundancy

This study is a continuation of the preceding one. The specific purpose was to compare the relationships between the abilities of deduction, verbal, reasoning, and perceptual speed, and comprehension of a reading passage written at different levels of redundancy. Comprehension was measured by a multiple-choice test over the material contained in the reading passage.

Method

Subjects. The subjects were seventy-three sixth-grade pupils from a predominantly Negro elementary school. All sixth-grade pupils were tested but only those who had relevant ability measures on record were included in the study. The ability tests had been given approximately eight months earlier. The median verbal IQ of the subjects was 78. The study group was chosen because their cognitive ability test data were available.

The subjects were randomly assigned to two treatment groups. After the data for those having ability scores were tabulated, it was found that 36 had received Treatment 1 and 37 had received Treatment 2.

Materials. The original and classmate versions of "The Shining Metal" (5,6) were chosen as experimental passages because they were short enough to be read in a class period and because the two versions had greater differences in redundancy than did the other pairs of passages. In the earlier study, sixth-grade pupils restored 37 percent of the deleted words to the classmate version and 21 percent to the original version of "The Shining Metal" passage. The classmate version contained a total of 897 words and the original 1,175 words. The stories were typed with single spacing and reproduced by mimeograph.

A thirty-item four-choice test was constructed on the content of "The Shining Metal." The items dealt with general and specific story content common to both versions. Kuder-Richardson 20 reliabilities were computed for the test, using the separate data from each of the treatment groups. The reliability coefficient was .84 for the classmate group and .79 for the original group. The achievement test is given in Appendix A.

Four cognitive abilities were used as predictor variables and these are described below.

A. Deduction - The ability to interpret correctly a verbal argument or progression of evidence and to form a valid inference based on the interpretation [Test 8, "Inferences," of the California Test of Mental Maturity, (40)].

B. Verbal - The ability to understand ideas expressed in words [The "Words Test" and "Pictures Test" of the Primary Mental Abilities, (43)].

C. Reasoning - The ability to solve logical problems [The "Figure Grouping" and "Word Grouping" tests from the Primary Mental Abilities (43)].

D. Perceptual Speed - The ability to recognize similarities and differences between objects or symbols with speed and accuracy [The "Perceptual Speed" test from the Primary Mental Abilities (43)].

Procedure. The reading passages were distributed to students with instructions to study them carefully because they would be asked questions about them later. They were instructed to record the number of minutes they spent studying the passage and to close their booklets when finished. After the pupils had studied the passage, the booklets were put aside and the achievement test was distributed. Time was allowed for the pupils to attempt all test items.

Analyses. The mean time spent on studying the passage and the mean achievement score were computed for the classmate and for the original version groups.

Because the number of words contained in the two versions was different, each achievement score was divided by the number of minutes devoted to studying the passage to derive a rate-of-learning or "efficiency" score.

The principal data were analyzed by plotting the regression line for achievement efficiency on ability under each of the treatment conditions. A separate plot of the two regression lines was made for each of the ability measures. The efficiency scores were used as the criterion of achievement in constructing the regression lines.

In addition to the regression plots, the mean, standard deviation, correlation with achievement, efficiency, regression coefficient, and regression line crossover point were computed for each ability measure. The differences between slopes of the paired regression lines were tested for significance using an adaptation of a technique described by Kenny and Keeping in Mathematics of Statistics (32).

Results

The group given the original version of the story devoted more time to the study of the passages than the group that studied the classmate version. The mean time for each is shown in Table 19. The difference between mean study times is not as large as might be expected on the basis that the classmate version contained only 76% as many words as the original.

TABLE 19

Mean Achievement Scores, Study Times, and Efficiency Scores for the Two Treatment Groups

	Treatment Group	
	Classmate	Original
Achievement Score	15.7	14.4
Study Time	15.6	16.1
Efficiency Score	1.1	1.0

The mean achievement scores were higher for the total group which studied the classmate version than for the group which studied the original (Table 19). The classmate version of the material written at a simplified level was perhaps more nearly at the educational level of the subjects than the original version.

The difference between overall means indicates that the classmate version led to greater achievement than the original version. However, this inference is probably valid only for groups similar to that on which the study was done. To determine if the increase in achievement attributed to the classmate version applies throughout the range of each ability, the regression lines for achievement on ability were plotted for each of the ability measures. These regression plots are shown in Figures 1 through 4. Differences between the slopes of the paired regression lines were significant only for reasoning ability.

Cronbach and Gleser (12) discuss the possible interpretations of such regression lines. Their discussion is in the context of personnel placement; however, the assignment of pupils to different forms of content presentation is an analogous use of the technique.

With two treatments, there are three possibilities:

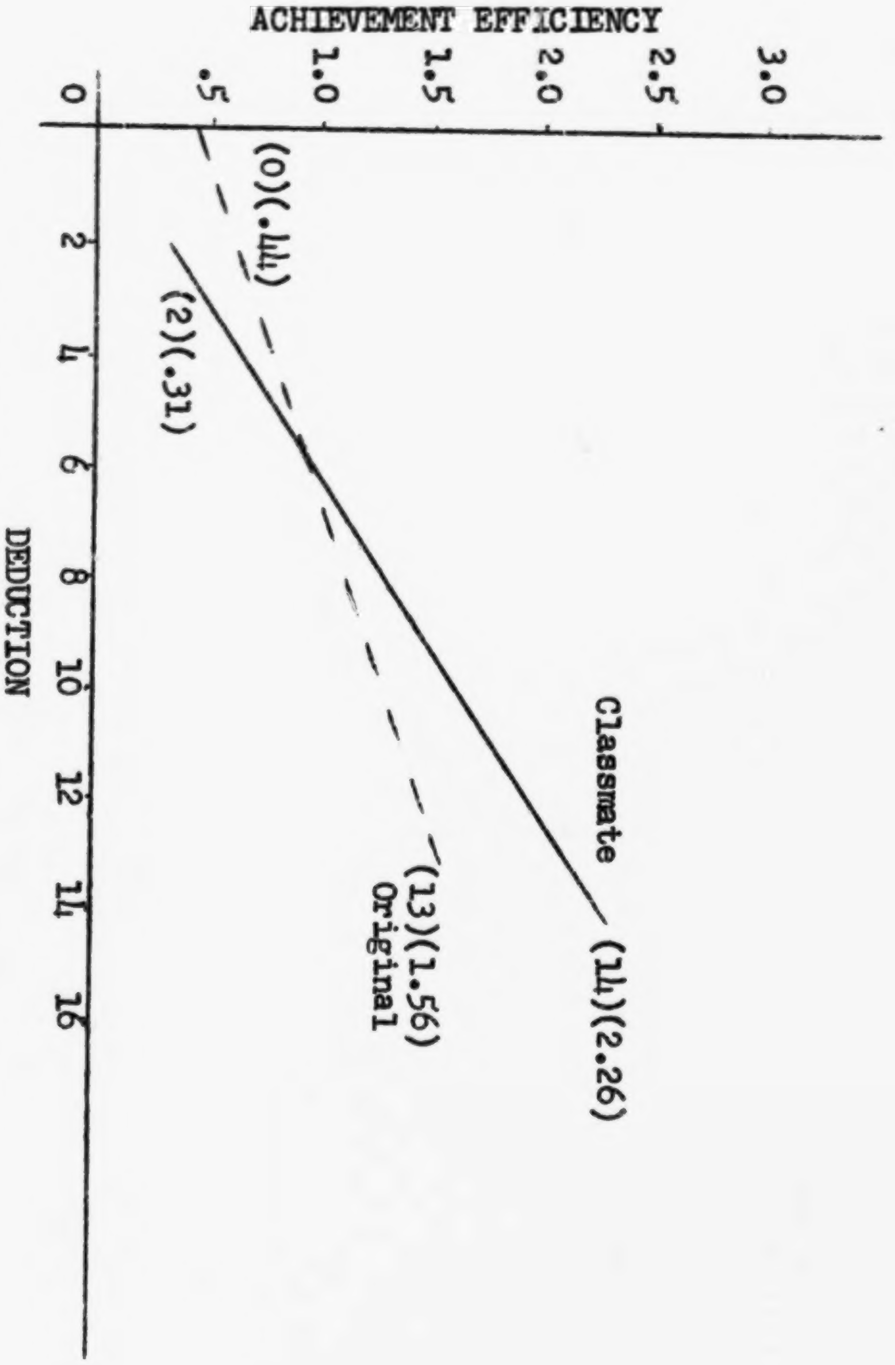

Figure 1. Regression lines of achievement efficiency on deduction ability for classmate and original study groups.

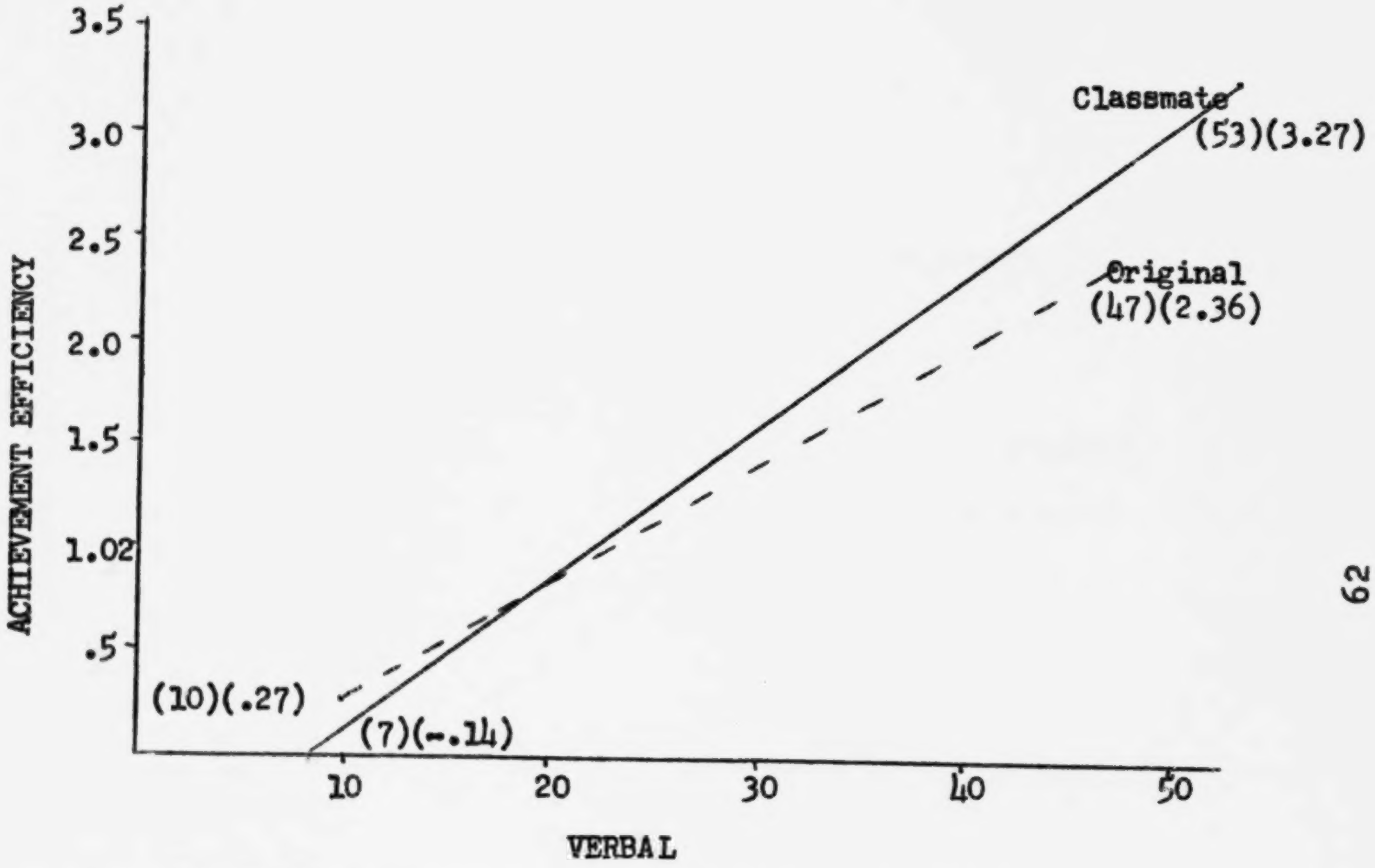

Figure 2. Regression lines of achievement efficiency on verbal ability for classmate and original study groups.

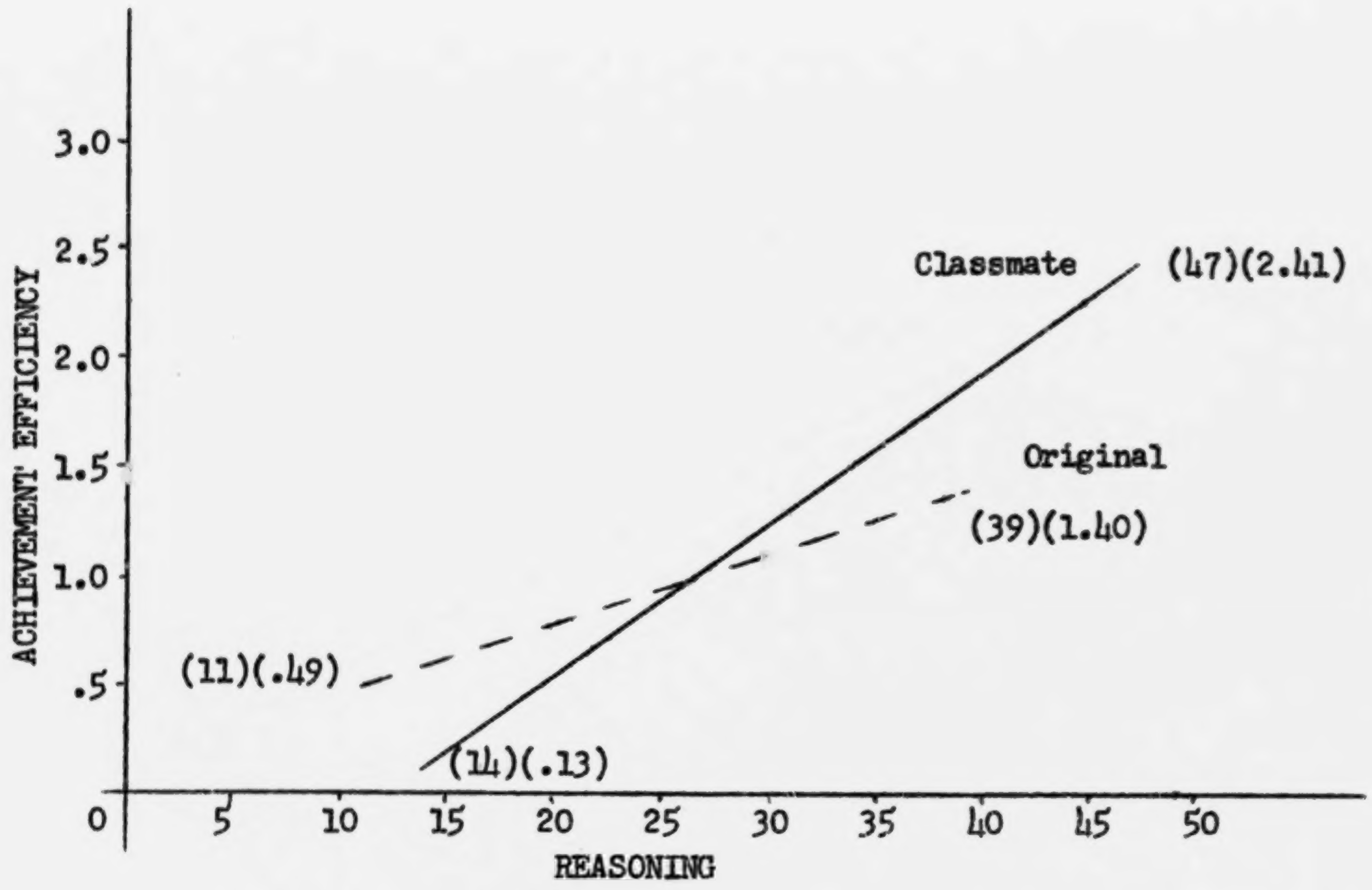

Figure 3. Regression lines of achievement efficiency on reasoning ability for classmate and original study groups.

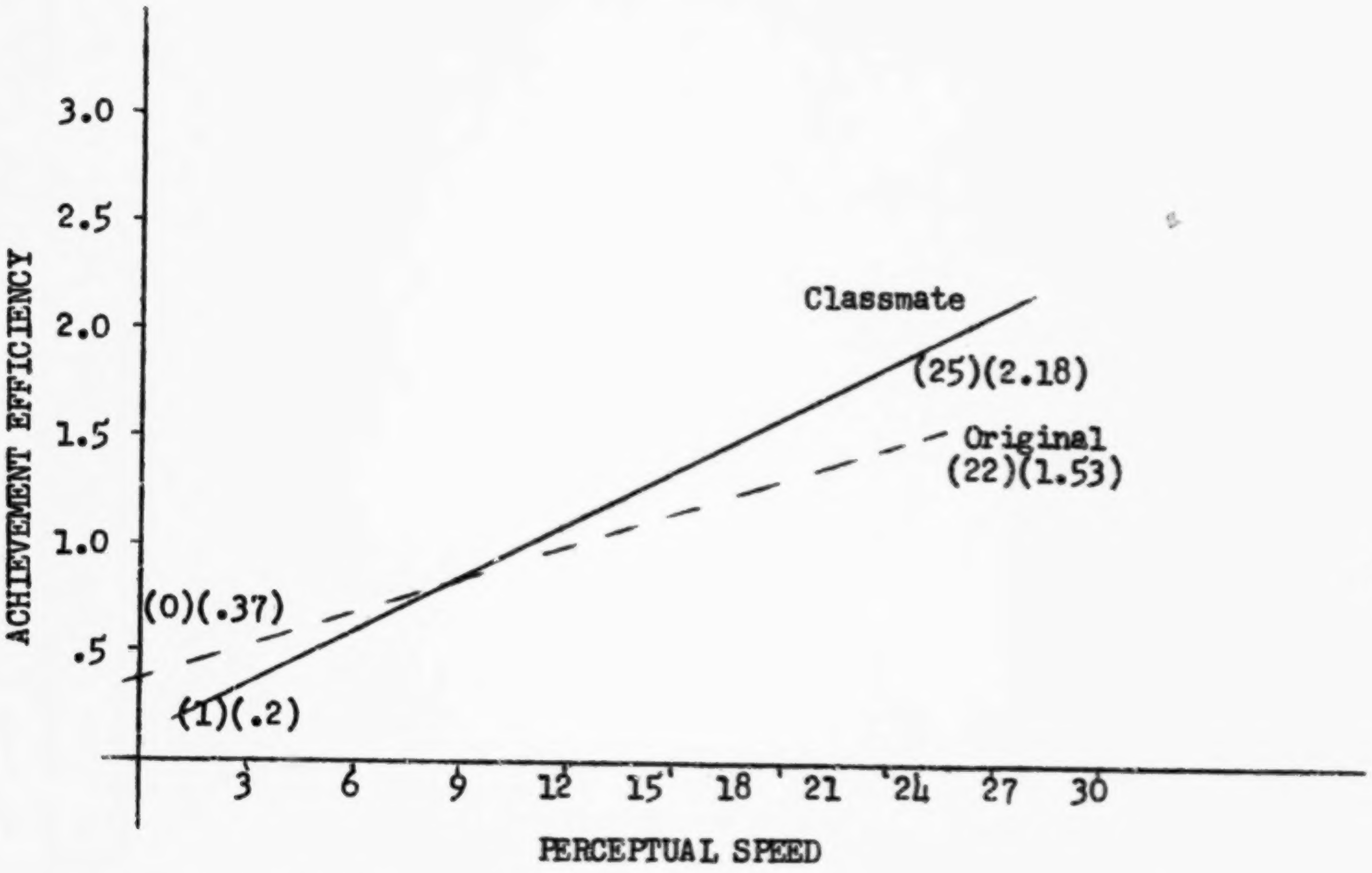

Figure 4. Regression lines of achievement efficiency on perceptual speed for classmate and original study groups.

A. The payoff functions (regression lines) are identical, in which case placement is valueless because the treatment to which a man is assigned makes no difference in payoff.

B. The payoff function for one treatment is uniformly higher than the other for all score levels. In this event it is unwise to divide the groups. All persons should be given the superior treatment unless institutional constraints require both treatments to be used (e.g., where there are too few therapists to give psychotherapy to all patients).

C. The payoff function intersects somewhere within the score range. If this is the case, it is profitable to assign men to different treatments, so that each receives the treatment that "gets the most out of him."

The regression lines for all four abilities conform to possibility C above (Figs. 1, 2, 3, 4). These lines reflect an interaction effect and cross within the score range. For all four ability measures the classmate version resulted in greater "payoff" in achievement at the upper ability levels and less "payoff" at lower ability levels. For example, a student having a reasoning score of 35 and studying the original version would have a predicted achievement (efficiency) score of 1.3 while the same subject would have a predicted achievement score of 1.6 if he studied the classmate version. In other words, a student having a reasoning score of 32 who studied the classmate version would have the same predicted achievement score as a student having a reasoning score of 37 who studied the original version.

Another comparison of within-treatment regression appears in Table 20. The regression slopes for deduction, reasoning, and perceptual speed differ more from those for verbal ability. The crossover points for all of these abilities fall below the mean achievement scores.

Although only the reasoning regression slopes were significantly different, there were consistencies across the four plots. The crossover points were generally to the left of the mean. This suggests that the differences in slope may have been created by

TABLE 20

Comparison of Within-Treatment Regressions

Ability Factor	Treatment Group	Mean	S.D.	Correlation With Criterion	Regression Coefficient	t	t for difference	Crossover Point Raw	Crossover Point Z
Deduction (CTMM)	Classmate	6.8	2.4	.55	.16	3.9		6.0	-0.3
	Original	6.4	2.7	.42	.09	2.8	1.4	6.0	-0.2
Verbal (PMA)	Classmate	23.6	8.0	.84	.07	9.0		18.0	-0.7
	Original	22.9	7.0	.72	.06	6.2	1.0	18.0	-0.7
Reasoning (PMA)	Classmate	28.0	7.1	.69	.07	5.6		27.0	-0.1
	Original	26.6	6.2	.37	.03	2.3	2.1	27.0	-0.1
Perceptual Speed (PMA)	Classmate	11.8	4.8	.56	.08	3.9		8.0	-0.8
	Original	11.9	4.5	.43	.05	2.8	1.1	8.0	-0.9
Achievement Score/Time	Classmate	1.1	.7						
	Original	1.0	.6						

achievement differences at the upper end of the ability measure range. It is possible that differences at the lower end of the ability range are meaningless.

In interpreting the above results, a linear relationship between each ability and achievement has been assumed. Where this assumption is not valid, the regression lines give a distorted picture. The studies should be repeated using groups of students having a full range of the relevant abilities, and the linearity assumption should be tested.

Discussion

The results of this study provide some support to the notion that a pupil's achievement depends upon an interaction between his aptitude and the redundancy level of the instructional material used. The data do not permit generalization about the form of this relationship throughout a wide range of these abilities.

E. Summary

The studies reported in this section deal with redundancy and the execution of them spanned nearly the entire period of the project. The first of them attempted to ascertain what elements of style are related to redundancy. It was felt that knowledge of these would enable the development of a set of materials, each differing in redundancy and all dealing with the same content, that would be useful in examining the ATI theory. The second study attempted to identify what cognitive abilities are most related to redundancy. The most highly related of these were logical reasoning, vocabulary and inference. It was felt that the results of these two studies would provide information about how to prepare instructional materials of known redundancy and to identify cognitive aptitudes or skills which would be most likely to interact with level of redundancy. The third study was of sets of published graded reading materials to determine if different versions of the same published work differed significantly in level of redundancy. The results generally supported the notion that a strong relationship exists between graded reading materials and their redundancy level and thereby supplied the investigators with rather polished reading passages which have high interest value and different and known levels of redundancy. These materials were then used in the fourth study which resulted in the identification of a rather pronounced interaction of reasoning ability and materials graded according to redundancy. This interaction appears to be sufficiently strong to enable assigning students on the basis of aptitude to different treatments in a manner which would optimize overall achievement.

V. ELEMENTARY SET CONCEPTS

A. Introduction

The purpose of preceding studies was to attempt to identify form-of-content variables in existing textual material. The purpose of the study reported here was to construct four sets of instructional materials, all dealing with the same content but differing in the aptitudes required by them for optimal achievement. The subject matter was elementary set concepts. The major aptitude contrasts were verbal versus figural, and inductive versus deductive reasoning.

One specific hypothesis resulting from the aptitude treatment interaction theory (ATI) was that verbal ability measures would have significantly higher regression coefficients when used to predict achievement on verbally presented materials than they would have when used to predict achievement on materials presented figurally (or vice versa). Similar hypotheses concerning the other ability contrasts were also inherent in the present study.

B. Method

Subjects.

The sample was drawn from two elementary schools. In one school, 222 students were tested and 204 were tested in the other school. Four fifth and four sixth grade classes were used from each school. Approximately equal numbers of males and females were included from each school and appeared in the total sample.

Learning Materials.

Four short sets of learning materials were developed to teach five basic concepts of mathematical sets. These were: definition of a set, definition of elements of sets, disjoint sets. Each of the four different types of material was from four to six pages in length and provided instruction on only the five concepts.

The verbal deductive (VD) material presented the five basic concepts in a direct expository manner devoid of any figural notation like Venn diagrams and intersection and union symbols. The style was deductive; a concept was presented and followed by several samples.

The verbal inductive (VI) material contained no figural or symbolic notation and its style was different from the (VD) materials. Numerous examples were given that were stated to be, say, disjoint sets. The student was to induce or discover the appropriate definitional attribute for the concept of disjoint sets. The actual concept was not explained or defined.

The figural deductive (FD) material minimized the verbal content and maximized usage of set notation and symbols. Venn diagrams were used whenever possible in the definitional and examples parts for each concept. A deductive style was used to present the defining attributes of the concept and then several appropriate examples. It should be stressed that the figural material (both FD and FI) minimized verbal usage, it did not eliminate it.

The figural inductive (FI) material minimized the verbal narrative in order to accentuate the symbolic, figural notation. The main difference between the FD and the FI type was the presentation style.

The four sets of materials were used in two pilot investigations. The major purposes of these were to determine if vocabulary changes were necessary, the best format for the achievement test, and the grade levels at which the materials would be appropriate. As a result, the materials which were finally used in the study underwent three revisions. Complete sets of the materials are shown in Appendix B.

Criterion and Ability Tests.

In conjunction with the four sets of material, a criterion "set" test was developed that contained 28 items. Items were constructed for the last four (elements, disjoint sets, intersection and union) of the five concepts in the instructional material. Items were constructed for each concept to represent both the verbal and figural dimensions; i.e., a question about the elements of the "set of all animals" would have items written in both verbal and figural modes. Of the total of twenty-eight items, twelve were verbal and sixteen were figural.

The battery of ability tests consisted of the Inference test (deductive reasoning) from the California Test of Mental Maturity, Level 2, (40), and five subtests from the Primary Mental Abilities, Grade Level 4-6, (43), tests. The verbal PMA tests were Words Test and Pictures Test. PMA Figure Grouping Test and Word Grouping Test were used as tests of inductive reasoning, and PMA Perceptual Speed Test was used as a test of figural ability.

Procedure

The study was conducted in two consecutive days at each school. Each day's testing required about 1 to 1 1/2 hours from each class. In every class, the first day was devoted solely to administering ability tests. The syllogisms test was given first and then relevant subtests from the PMA.

The first task given on the second day was the learning materials. The four sets of materials were distributed in a 1,2,3,4,1,2,3,4 etc., order throughout the class. Therefore each class had a somewhat different distribution of the VD, VI, FD, and FI materials. The overall distribution of these materials appears in Table 21.

Before the learning materials were distributed, the administrator in charge wrote each of the five basic concept words on the board and pronounced them. He explained that these words were the concepts to be learned. After being given the materials, students were instructed to read through them carefully because they would be tested on their contents after finishing.

TABLE 21

Number of Students Who Received Each Kind of Material

	VD	VI	FD	FI	Total
School 1	54	51	56	61	222
School 2	54	49	51	50	204
Total	108	100	107	111	426

All students began to read the materials at the same time. When each finished he exchanged his booklet for a criterion achievement test.

Very little help was given to students in pronouncing words or explaining the concepts. They were told to do the best they could and, if they did not understand one concept, then to go on to the next. This procedure was also followed when they took the achievement test.

C. Results

The means and standard deviations for the verbal, figural and total criterion measures for each of the four treatment groups appear in Table 22. Means and

TABLE 22

Means and Standard Deviations of Three Criterion Measures for Four Treatment Groups

Criterion Test	Treatment Group							
	VI		VD		FI		FD	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Verbal	5.91	2.17	5.49	2.11	5.85	2.22	5.59	2.11
Figural	5.41	2.04	5.06	2.20	5.12	2.19	5.22	2.13
Total	12.56	3.89	11.93	4.12	12.19	4.23	11.98	4.16

standard deviations of the ability measures for the four treatment groups appear in Table 23. Matrices of

TABLE 23

Means and Standard Deviations of Ability Measures for Four Treatment Groups

Ability Test	Treatment Group							
	VI		VD		FI		FD	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Inference	7.57	2.72	7.75	2.62	7.49	2.90	7.33	3.02
Words Test	14.68	5.83	14.79	5.75	14.33	5.67	13.99	5.97
Pictures Test	12.59	4.06	11.74	3.89	12.39	4.30	12.56	4.78
Figure Grouping Test	15.78	3.75	15.68	4.07	15.46	3.71	16.60	6.77
Word Grouping Test	14.69	4.18	14.60	6.89	14.11	4.04	14.00	4.31
Perceptual Speed Test	15.18	6.08	13.99	5.90	14.23	6.35	14.53	5.58

intercorrelations of the ability and criterion variables for each treatment group appear in Tables 24 through 27.

Visual inspections of the means, standard deviations, and intercorrelations of the ability measures indicate no differences large enough to preclude comparing the regression coefficients for the four groups.

TABLE 24

Intercorrelations of Ability and Criterion Measures for Treatment Group VI

	1	2	3	4	5	6	7	8	9
Verbal Criterion	1.00	.50	.87	.22	.44	.36	.28	.39	.47
Figural Criterion		1.00	.85	.23	.28	.32	.24	.32	.26
Total Criterion			1.00	.29	.43	.40	.30	.41	.44
Inference				1.00	.52	.42	.28	.50	.29
Words Test					1.00	.57	.21	.72	.49
Pictures Test						1.00	.20	.47	.38
Figure Grouping Test							1.00	.31	.31
Word Grouping Test								1.00	.49
Perceptual Speed Test									1.00

TABLE 25

Intercorrelations of Ability and Criterion Measures for Treatment Group VD

	1	2	3	4	5	6	7	8	9
Verbal Criterion	1.00	.49	.83	.37	.49	.46	.17	.29	.42
Figural Criterion		1.00	.87	.35	.42	.48	.31	.29	.33
Total Criterion			1.00	.42	.53	.55	.28	.34	.44
Inference				1.00	.60	.57	.40	.45	.33
Words Test					1.00	.64	.36	.55	.51
Pictures Test						1.00	.34	.43	.35
Figure Grouping Test							1.00	.51	.26
Word Grouping Test								1.00	.31
Perceptual Speed Test									1.00

TABLE 26

Intercorrelations of Ability and Criterion Measures for Treatment Group FI

	1	2	3	4	5	6	7	8	9
Verbal Criterion	1.00	.50	.85	.14	.46	.40	.26	.39	.24
Figural Criterion		1.00	.85	.24	.44	.40	.17	.44	.24
Total Criterion			1.00	.22	.53	.48	.28	.48	.29
Inference				1.00	.46	.40	.23	.44	.36
Words Test					1.00	.67	.26	.57	.40
Pictures Test						1.00	.25	.50	.36
Figure Grouping Test							1.00	.40	.21
Word Grouping Test								1.00	.26
Perceptual Speed Test									1.00

TABLE 27

Intercorrelations of Ability and Criterion Measures for Treatment Group FD

	1	2	3	4	5	6	7	8	9
Verbal Criterion	1.00	.53	.85	.29	.41	.46	.02	.35	.45
Figural Criterion		1.00	.87	.31	.41	.50	.06	.44	.45
Total Criterion			1.00	.34	.48	.57	.05	.46	.54
Inference				1.00	.53	.53	.11	.50	.53
Words Test					1.00	.56	.09	.64	.52
Pictures Test						1.00	.10	.46	.51
Figure Grouping Test							1.00	.16	.16
Word Grouping Test								1.00	.52
Perceptual Speed Test									1.00

The multiple regression coefficients and multiple correlations for each of the criterion measures for each of the four groups appear in Tables 28 through 30.

TABLE 28

Multiple Regression Coefficients and squared Multiple Correlations of Ability Measures and the Verbal Criterion For the Four Treatment Groups.

	Group			
Ability	VI	VD	FI	FD
Inferences	-.065	.053	-.129	-.067
Words Test	.093	.073	.118	.046
Pictures Test	.058	.129	.073	.122
Figure Grouping Test	.079	-.042	.061	-.021
Word Grouping Test	.008	.007	.089	.031
Perceptual Speed Test	.099	.081	.022	.102
Intercept	1.434	1.920	1.711	2.327
R^2	.307	.321	.276	.297

TABLE 29

Multiple Regression Coefficients and Squared Multiple Correlations of Ability Measures and the Figural Criterion for the Four Treatment Groups.

	Group			
Ability	VI	VD	FI	FD
Inference	.017	.020	-.029	-.065
Words Test	.000	.030	.075	.008
Pictures Test	.095	.189	.070	.151
Figure Grouping Test	.068	.070	-.014	-.009
Word Grouping Test	.067	-.004	.146	.108
Perceptual Speed Test	.027	.051	.020	.080
Intercept	1.628	.504	1.257	1.190
R^2	.160	.283	.259	.339

TABLE 30

Multiple Regression Coefficients and Squared Multiple Correlations of Ability Measures and the Total Criterion for the Four Treatment Groups.

	Group			
Ability	VI	VD	FI	FD
Inference	.010	.076	-.183	-.195
Words Test	.097	.116	.216	.062
Pictures Test	.165	.350	.166	.325
Figure Grouping Test	.147	.030	.089	-.029
Word Grouping Test	.058	.004	.244	.157
Perceptual Speed Test	.146	.151	.053	.223
Intercept	3.612	2.883	2.839	3.517
R^2	.302	.391	.362	.441

Although the regression equations in Tables 28 through 30 provide some evidence of the existence of aptitude treatment interactions, the results are not clear-cut. It was felt that the three sets of four regression equations, each including six predictor variables were based on interrelationships which were too complex to yield easily interpretable results. Therefore, to investigate the aptitude treatment interaction hypothesis, treatment groups were combined into verbal and figural groups, and then into deductive and inductive groups, and only the relevant ability measures in each instance were used. This reduced the number of variables and equations which had to be considered simultaneously. Also, it enables basing each regression equation on a much larger sample, thereby increasing its stability. The means and standard deviations of the combined groups appear in Table 31 and the correlation matrices appear in Tables 32 through 35.

TABLE 31

Means and Standard Deviations of Criterion and Ability Measures for Groups V, F, D and I

Criterion Test	Group							
	V (N=208)		F (N=218)		D(N=215)		I (N=211)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Verbal	5.70	2.15	5.72	2.18	5.54	2.12	5.88	2.20
Figural	5.23	2.13	5.17	2.17	5.14	2.17	5.26	2.13
Total	12.23	4.03	12.09	4.21	11.95	4.15	12.37	4.09
Inference	7.66	2.68	7.41	2.97	7.54	2.84	7.53	2.82
Words Test	14.73	5.81	14.16	5.84	14.39	5.89	14.50	5.77
Pictures Test	12.14	4.00	12.47	4.56	12.15	4.39	12.48	4.20
Figure Grouping Test	15.73	3.93	16.02	5.98	16.14	5.61	15.62	3.74
Word Grouping Test	14.64	5.75	14.05	4.18	14.30	5.76	14.38	4.13
Perceptual Speed Test	14.57	6.03	14.38	5.99	14.26	5.76	14.69	6.25

TABLE 32

Intercorrelations of Ability and Criterion Measures for Group V

	1	2	3	4	5	6
Words Test	1.00	.60	.50	.47	.35	.48
Pictures Test		1.00	.37	.42	.41	.48
Perceptual Speed Test			1.00	.45	.30	.44
Verbal Criterion				1.00	.50	.85
Figural Criterion					1.00	.86
Total Criterion						1.00

TABLE 33

Intercorrelations of Ability and Criterion Measures for Group F

	1	2	3	4	5	6
Words Test	1.00	.61	.46	.44	.42	.51
Pictures Test		1.00	.43	.43	.45	.52
Perceptual Speed Test			1.00	.33	.33	.40
Verbal Criterion				1.00	.51	.85
Figural Criterion					1.00	.86
Total Criterion						1.00

TABLE 34

Intercorrelations of Ability and Criterion Measures for Group D

	1	2	3	4	5	6
Inference	1.00	.20	.45	.32	.33	.37
Figure Grouping Test		1.00	.29	.07	.16	.14
Word Grouping Test			1.00	.31	.34	.38
Verbal Criterion				1.00	.51	.84
Figural Criterion					1.00	.87
Total Criterion						1.00

TABLE 35

Intercorrelations of Ability and Criterion Measures for Group I

	1	2	3	4	5	6
Inference	1.00	.25	.47	.18	.24	.25
Figure Grouping Test		1.00	.36	.27	.21	.29
Word Grouping Test			1.00	.39	.38	.45
Verbal Criterion				1.00	.50	.87
Figural Criterion					1.00	.85
Total Criterion						1.00

The regression equations and squared multiple correlations for the three criterion measures appear in Table 36 for groups V and F and in Table 37 for groups D and I.

Evaluating the presence of the hypothesized aptitude treatment interaction was done by noting the significance of the difference between pairs of regression coefficients (32). Each pair of coefficients involved the same ability and criterion measures for the two groups being considered. Because the patterns of coefficients for all criterion measures were highly similar for each group, only coefficients involving the total criterion measure were tested. The results of the statistical analyses appear in Table 38.

TABLE 36

Regression Equations and Squared Multiple Correlations For Three Criterion Measures for Groups V and F.

Ability Measures	Verbal Criterion		Figural Criterion		Total Criterion	
	V	F	V	F	V	F
Words Test	.081	.090	.038	.072	.133	.177
Pictures Test	.099	.109	.157	.134	.277	.281
Perceptual Speed Test	.097	.044	.051	.045	.164	.111
Intercept	1.886	2.441	2.033	1.834	4.508	4.469
R^2	.301	.244	.200	.251	.334	.347

TABLE 37

Regression Equations and Squared Multiple Correlations For Three Criterion Measures for Groups D and I.

Ability Measures	Verbal Criterion		Figural Criterion		Total Criterion	
	D	I	D	I	D	I
Inference	.176	-.015	.164	.048	.371	.060
Figure Grouping Test	-.014	.088	.018	.040	.006	.156
Word Grouping Test	.077	.183	.086	.169	.188	.373
Intercept	3.339	1.986	2.384	1.828	6.377	4.109
R^2	.138	.169	.155	.156	.194	.220

TABLE 38

Tests of Significance Between Pairs of Regression Coefficients for Groups V and F and Groups D and I.

Group V Versus Group F			Group D Versus Group I		
Ability	t-ratio	P	Ability	t-ratio	P
Words Test	-.59	>.05	Inference	2.94	<.05
Pictures Test	-.04	>.05	Figure Grouping Test	-1.68	>.05
Perceptual Speed Test	.85	>.05	Word Grouping Test	-2.80	<.05

None of the t-ratios for the V-F comparisons is significant, so there is no support for aptitude treatment interaction. However, the D-I contrasts support the hypothesis because two of the t-ratios are significant at the .05 level and the differences in both cases are in the hypothesized direction. Thus the Inference Test (deduction) is a better predictor for the D materials than for the I materials. For the Word Grouping Test (induction) the converse is true.

D. Discussion

The contrast of the groups formed by merging all students who were exposed to deduction and induction materials, respectively, demonstrates the feasibility of constructing parallel sets of instructional materials that require different abilities for optimal achievement. The failure of the verbal-figural contrast to show ATI effects could have occured because variations along this dimension are unimportant in achievement of elementary set concepts. On the other hand, it is possible that the hypothesized effects failed to appear because the verbal-figural contrasts were not sufficiently distinct and/or because the figural ability measures were not appropriate.

E. Summary

The purpose of this study was to construct four sets of instructional materials, all dealing with the same elementary set concepts but differing in the aptitudes required by them for optimal achievement. Each set of materials was written to emphasize the use by students of one of the following ability combinations: Verbal-deductive, verbal-inductive, figural-deductive, figural-inductive.

Subjects for the study were 426 fifth and sixth grade students. Approximately equal numbers of males and females were represented in the sample. Within each classroom approximately equal fourths of the students received each of the four sets of material.

Two criterion measures, one containing figural items and the other containing verbal items, were administered. In addition all students were administered appropriate tests of the abilities under study.

Regression equations of the criterion measures on the ability measures were computed for each of the four groups. While some evidence of aptitude treatment interactions (ATI) were noted they were not as hypothesized and were not easily interpretable. In order to simplify the analyses the treatment groups were combined into verbal and figural groups and then into deductive and inductive groups. Multiple regression equations for the criterion measures were computed for each pair of groups using only the relevant ability measures. The verbal and figural group equations gave no indication of ATI. The contrast of inductive and deductive group equations did support the ATI theory. The deductive test was a significantly better predictor of achievement of the deductive materials whereas the verbal inductive test was significantly better for predicting the inductive materials. The figural inductive test did not predict achievement well for either set of materials. The results of the study demonstrate, at least partially the feasibility of constructing parallel sets of material that require different abilities for optimal achievement.

It is possible that the failure of the verbal-figural contrast to yield different regression equations was due to insufficient differences between the sets of materials. Rewriting of the materials to drive them apart on the verbal-figural dimension might allow future experiments to demonstrate differential predictions.

VI. STUDENT AND TEACHER ABILITY PATTERNS AND ACHIEVEMENT IN CHEM-STUDY*

A. Introduction

The research presented in this section of the report is different in that the treatments described here are not sets of instructional materials which have different levels of some form-of-content variable but are, instead, teachers having different ability patterns. Reviews of investigations of teaching methods (44) and of teacher characteristics (27) indicate that most of the reported studies deal with evaluating methods or teacher characteristics without regard for individual differences that exist among students.

The aptitude treatment interaction hypothesis clearly applies to relationships between sets of learning materials and student abilities. Although it is not as evident, the hypothesis also applies to relationships between teacher abilities and student abilities. It seems reasonable to believe that a teacher of statistics who has high spatial ability would tend to capitalize on this ability in his lectures and explanations. A student also high in spatial ability might tend to profit more from this type of instruction than a student with less spatial ability.

The specific purpose of this study was to determine whether achievement of students taking a course in CHEM-STUDY high school chemistry is greater for those who are similar in certain cognitive abilities to their teacher than it is for students who are less similar to their teacher when level of general ability is held constant.

---

*Cleare, B. E. An Investigation of the Interaction between Student-Teacher Cognitive Ability Patterns Using Achievement in the Chemical Education Material Study Chemistry Course as the Criterion Variable. Unpublished doctoral dissertation, Florida State University, 1966.

B. Method

Subjects

The teachers who participated in the study were 13 of 18 teachers who were enrolled in a CHEM-STUDY institute and who were also teaching CHEM-STUDY to high school students. The students involved were 917 10th grade CHEM-STUDY students who were taught by the 13 teachers. The number of students taught by any one teacher ranged from 21 to 112.

Tests

The battery of ability measures consisted of seven tests from the ETS Kit of Reference Tests (22). Previous unpublished investigations indicate that the tests had relatively low intercorrelations and were of adequate reliability when used with high school students. The seven tests and the factors they purport to measure follow:

A. Wide Range Vocabulary Test (V-3), verbal comprehension.

B. Inference Test (Rs-3), syllogistic reasoning.

C. Necessary Arithmetic Operations (R-4), general reasoning.

D. Paper Folding Test (Vz-2), visualization.

E. Gestalt Transformation (Re-1), semantic redefinition.

F. Hidden Figures Test (Cf-1), flexibility of closure.

G. Locations Test (I-2), induction.

The criterion measures were the standardized CHEM-STUDY unit tests 1, 2, and 3 and the mid-semester examination. A pretest was made from the mid-semester examination by randomly selecting 25 of the 50 items.

Procedure

The ability measures were administered to the 13 teachers during a meeting of the Institute prior to the beginning of the school year. Ability test and pretest materials and detailed instructions for their administration were given to the teachers, and they administered the tests to their students during the first month of school. The unit and mid-semester examinations were administered by the teachers at appropriate times.

Analysis

The raw ability scores of the 917 students were converted to T scores, and the teachers' scores were then converted to T scores on the basis of the student norms. A random sample of 200 students was selected, and reliabilities for six of the tests were estimated from the correlations of part-scores corrected for length by the Spearman-Brown prophecy formula. Only one part of the Wide Range Vocabulary Test was given, so its uncorrected reliability was estimated from previously obtained data from approximately 200 10th grade students.

Coefficients of cognitive pattern and level similarity were computed for each student and his teacher using the intra-class correlation technique presented by Haggard (29). The 917 students were ranked on the coefficient of pattern similarity and were divided into approximately equal thirds. Each subgroup was subdivided into two level categories, and each of these groups was divided into males and females. The number of students in each of the various subgroups is shown below where L_1 and L_2 indicate high and low levels and P_1, P_2, and P_3 indicate high, medium, and low pattern similarity respectively.

	L_1		L_2	
	M	F	M	F
P_1	83	97	72	54
P_2	100	53	84	72
P_3	91	37	115	61

The dependent variable was the number of correct pretest items subtracted from the total number of items that each student answered correctly in all unit tests and the mid-term examination, divided by the total number of items he was given. The adjustment for total items was necessary because two teachers failed to administer one unit test. Factorial analysis of variance and covariance procedures were used to determine differences in mean achievement among the various subgroups. Since the design was not orthogonal the general linear hypothesis was used in the actual analysis.

Simple regression coefficients for the regression of the achievement measure on each ability measure were computed for each of six teachers who taught more than ninety students. A visual inspection was made of the regression coefficients for each ability to determine whether they covaried with the teacher's scores on the ability.

C. Results

The reliabilities of the seven cognitive tests computed from 200 subjects appear in Table 39. The intercorrelations of the tests computed from the data for all 917 subjects appear in Table 40.

TABLE 39

Reliabilities of Cognitive Tests

Wide Range Vocabulary Test (V-3)	.82*
Inference Test (Rs-3)	.60
Necessary Arithmetic Operations (R-4)	.74
Paper Folding Test (Vz-2)	.71
Gestalt Transformation (Re-1)	.62
Hidden Figures Test (Cf-1)	.73
Locations Test (I-2)	.69

*estimated from previously gathered data

TABLE 40

Intercorrelations of Cognitive Tests

	Rs-3	R-4	Vz-2	Re-1	Cf-1	I-2
V-3	.52	.42	.24	.32	.18	.26
Rs-3		.50	.29	.25	.24	.31
R-4			.40	.37	.34	.34
Vz-2				.32	.40	.30
Re-1					.22	.20
Cf-1						.31

The relationships between the magnitude of reliabilities of the tests and the magnitudes of their intercorrelations seem to warrant the interpretation that the tests are measures of separate cognitive abilities and that they are sufficiently precise for exploratory research.

The results of the 2 x 3 x 3 analysis of variance in which the dependence of CHEM-STUDY achievement on differences in sex, pattern, and level of teacher-student similarity was studied appear in Table 41. The means of the subgroups appear in Table 42. The analysis of variance results and the means support the hypothesis that teacher-student cognitive similarity is related to student achievement. The overall means for the three pattern groups were .52, .49 and .43 for high to low similarity respectively. The means were .54 and .42 for the high and low level similarity groups. Coefficients of correlation between pattern similarity coefficients and student ability scores were computed to determine whether the magnitude of the pattern similarity coefficients was independent of the magnitude of the students ability measures. The correlations were as follows: V-3, .59; Rs-3, .24; R-4, .11; Vz-2, -.18; Re-1, .11; Cf-1, -.22; I-2, -.07. Because all of these correlations are significant, it is possible that the significant F-ratio for pattern in the analysis of variance could be due simply to the ability scores of the students and not to the teacher-student similarity. The analysis of variance was repeated with the seven ability measures as covariates, and it was found that the pattern similarity effect was still highly significant

TABLE 41

Analysis of Variance of CHEM-STUDY Achievement by Sex, Level and Pattern*

Source of Variation	D.F.	Sums of Squares	F-Ratio	ρ
Sex (S)	1	149,984	8.25	<.01
Level (L)	1	2,860,162	157.30	<.01
Pattern (P)	2	1,116,722	30.71	<.01
S x L	1	10,378	.56	>.01
S x P	2	64,434	1.77	>.05
L x P	2	51,038	1.40	>.05
S x L x P	2	54,374	1.49	>.05
Error	905	16,455,556		
Total	916	21,450,636		

*In this analysis each student's score on the dependent variable was multiplied by 100.

TABLE 42

Means of CHEM-STUDY Achievement Scores by Sex, Level and Pattern

	Level 1		Level 2	
Pattern	Males	Females	Males	Females
P_1	.58	.50	.50	.44
P_2	.56	.43	.43	.42
P_3	.52	.40	.40	.37

(F = 16.95). Interestingly, the F-ratio for the level similarity effect fell to 7.60 but was still significant.

Table 43 shows the simple regression coefficients for each ability and CHEM-STUDY achievement for the students of each of six teachers. The standard scores for each teacher for each variable are given in Table 44. Table 45 contains the means and standard deviations of the achievement variable for the students of each teacher.

D. Discussion

Visual inspection of the regression coefficients and the teachers' scores does not reveal the relationships demanded by the similarity hypothesis. It is possible that one-to-one similarity in student-teacher ability is not the most relevant similarity. It might be that certain teacher abilities or characteristics enhance learning in students who are high in complementary abilities. An example might be a situation in which achievement is facilitated by high auditory decoding ability in students when their teachers are high in verbal fluency. When teachers have low verbal fluency, auditory decoding speed in students might have no relationship to achievement.

The relationship between the pattern coefficient and CHEM-STUDY achievement when student abilities are held constant does indicate the desirability for further research on this topic. Teacher variables other than cognitive abilities might be found to interact with student ability measures in affecting achievement.

E. Summary

The purpose of this study was to determine whether achievement of students of CHEM-STUDY high school chemistry is greater for those who are similar in cognitive abilities to their teachers than for students who are less similar to their teachers when level of general ability is held constant.

Thirteen chemistry teachers in a CHEM-STUDY Institute were administered seven relatively factorially pure ability measures. The teachers then

TABLE 43

Intercepts and Simple Regression Coefficients of Abilities for CHEM-STUDY Achievement

Ability		Teachers					
		T_1	T_2	T_3	T_4	T_5	T_6
Vocabulary	Intercept	.413	.353	.372	.494	.305	.244
	Coefficient	.012	.012	.013	.010	.015	.014
Inference	Intercept	.384	.282	.261	.427	.245	.165
	Coefficient	.013	.015	.019	.013	.017	.018
Necessary Arithmetic Operations	Intercept	.395	.248	.259	.388	.180	.148
	Coefficient	.009	.012	.014	.012	.016	.014
Paper Folding	Intercept	.384	.355	.328	.333	.339	.097
	Coefficient	.013	.009	.015	.020	.007	.021
Gestalt Transformation	Intercept	.495	.376	.412	.595	.347	.281
	Coefficient	.006	.012	.010	.001	.011	.010
Hidden Figures	Intercept	.481	.399	.427	.539	.400	.271
	Coefficient	.006	.006	.008	.005	.004	.008
Locations	Intercept	.475	.358	.408	.546	.351	.263
	Coefficient	.008	.011	.011	.007	.012	.014

TABLE 44

Teachers' Standard Ability Scores

Ability	Teachers T_1	T_2	T_3	T_4	T_5	T_6
Vocabulary	70	77	70	68	77	75
Inference	56	69	61	69	48	58
Necessary Arithmetic Operations	54	66	60	62	54	66
Paper Folding	50	57	45	57	55	55
Gestalt Transformation	50	54	54	64	67	72
Hidden Figures	48	53	50	54	54	41
Locations	53	57	61	53	51	42

TABLE 45

Means, Standard Deviations, and Numbers of Students of Each of Six Teachers on CHEM-STUDY Achievement

	Teachers T_1	T_2	T_3	T_4	T_5	T_6
Mean	.537	.465	.493	.608	.435	.349
Standard Deviation	.120	.131	.132	.132	.128	.153
N	.104	94	.111	95	96	99

administered the same tests to their students. Approximately 920 students were tested. Pattern similarity coefficients were computed between each teacher and each of his students. Male and female students were placed into three approximately equal pattern similarity groups (highly similar to teacher, similar to teacher and dissimilar to teacher). Each of these six groups was subdivided into two pattern level groups (about the same as the teacher and lower than the teacher). The criterion measure was performance on CHEM-STUDY achievement tests. The general linear hypothesis was used to analyze the data. It was found that all three main effects and no interactions were significant when student ability was held constant. Males had better achievement than females, students similar in level to the teacher did better than students lower than the teacher, and the greater the similarity in teacher-student pattern the greater the achievement.

Simple regression coefficients for each ability and achievement were computed separately for the students of each of six teachers. No set of regression coefficients for a particular ability appeared to be related to the teachers' scores for that variable.

The results of the study support the deduction from the ATI theory that students who are similar to their teachers in cognitive ability patterns achieve at a higher level than do students who are less similar to their teachers. The importance of a one-to-one similarity in student-teacher abilities could not be demonstrated.

VII. CONCEPT ATTAINMENT FROM VERBAL AND FIGURAL INSTANCES

A. Introduction

In the studies reported in prior sections, the form of content variables were embedded in textbooks or specially prepared instructional materials. Although as much control was exercised as possible over the student's prior knowledge of content when it seemed desirable to do so, the possibility exists that the results might have been somewhat effected by unequal and uncontrolled initial knowledge. As a consequence, the investigators chose to undertake a series of studies based on a standard task from experimental psychology about which students would be totally unfamiliar.

The task was concept attainment (7) where the relevant materials were either verbal or figural. Ramsey (38) used such materials and found that mastering verbal concepts required significantly more time than did mastering figural concepts, but that no other differences existed between the two forms in which the concept instances were presented.

The aptitude treatment interaction theory leads to the prediction of (a) no differences between forms (materials, content) when heterogeneous subject groups are compared, and (b) significant differences between the treatments when subjects are classified according to aptitude patterns which are believed to be relevant to the tasks.

The purpose of the study was to identify aptitude variables which might be differentially related to attainment of figural or verbal concepts and then to determine whether the hypothesized relationships could be empirically demonstrated.

B. Method

Task Analysis

In earlier project studies, the aptitudinal frame of reference was the Primary Mental Abilities. On the basis of the results of these studies, the experimenters felt that a broader conception of aptitudes was necessary to embrace the rich variety of apparent form of content variables. The Structure of Intellect (SI) model developed by Guilford (28) was chosen because it seemed to be easily related to the form of content variables. A task could be analyzed in terms of the SI model and the aptitudes relevant for its mastery would appear to be immediately known. This procedure was used in the analysis of the concept attainment task.

The concept attainment task required presenting the subject with an array of 64 stimulus cards. Six two-valued attributes constituted the array. The attributes were (a) number of borders (one or two), (b) kind of borders (broken or solid) and (c) number of figures (one or two), (d) their size (large or small), (e) their color (black or white) and (f) their shape (circle or ellipse). One of the 64 cards was presented to the subject as a "focus" card and he was told to discover the concept of which it was a member. He selected other cards, one-at-a-time, and was told by the examiner whether it was an example of the concept. The procedure continued until the subject formed an hypothesis about the concept and offered it to the examiner. If the correct concept was identified, card selection was terminated; if not, it continued until the correct concept was attained. Examples of a concept could be "black, circular figures" or "two borders, small figures." The subject was required to learn four concepts.

An analysis of the learning task in terms of the SI model suggested that learning the first concept required two major abilities. One of these is the ability to "cognize figural units" (CFU), which involves selecting a relevant detail out of a complex stimulus configuration. The other necessary ability is the "cognition of figural classes" (CFC). Probably the best-suited and available test for this ability is

Figure Classification, which involves defining classes of figures and assigning other figures to the correct classes.

When the learning task was altered so that verbal descriptions of the attributes replaced the actual values of the attributes, then a corresponding change of abilities was hypothesized to occur. The hypothesized abilities would become "cognition of symbolic units" (CSU) and "cognition of semantic classes" (CMC), respectively. The best tests of these abilities probably are Verbal Comprehension or Four Letter Words and Verbal Classification.

It was hypothesized that after the first concept had been learned, the "cognition of units" ability would become increasingly unimportant.

Production factors were not hypothesized in connection with this learning task because the dependent variables were number of cards chosen and number of hypotheses offered until the concept was attained. If the subjects were required to use the concept subsequently, then production factors might have been of importance.

Computer programs were used to administer the tasks to all subjects in the pilot study. In the major study some subjects received instruction from a human experimenter.

Tests

The tests which were actually used did not completely conform to the specifications derived from the task analysis. Hidden Patterns Test (CFU) and Figure Classification (CFC) were specified by the task analysis. Four Letter Words was presumed by the investigators to be a measure of CSU. No test of CMC was available at the time of the study so Letter Sets Test, which is probably a measure of CSC, was used. All of these tests were drawn from the ETS Kit of Reference Tests (22). Four Letter Words appears only in an earlier version of the KIT.

Subjects

Sixteen college sophomores were subjects in a pilot study which was conducted to test the experimental procedures and to determine the feasibility of subsequently using the Four Letter Words and Hidden Patterns Test in the major study. The other two tests were not administered to them. These students were from an original group of 80 who were reduced to 40 by excluding those who fell within one standard deviation of the mean on either test. The survivors were grouped into high-high, high-low, low-high and low-low test categories. Difficulties with the computer program or with the equipment itself further reduced the number of students with usable data to four per category. Each of these four groups was divided equally, one subgroup receiving the figural task and the other subgroup receiving the verbal task.

In the major portion of the study the four tests were administered to 96 senior and graduate students. Seventy-two of them were individually administered one of the two forms of the concept attainment task. Only 51 students actually completed the concept materials, 24 in the verbal concept group and 27 in the figural concept group.

Testing Conditions

The 64 concept attainment instances were drawn or typed on 3 x 3 inch cards and mounted in random order on square pieces of 1/4" plyboard. An instruction booklet was prepared and a computer-assisted instruction (CAI) program was written in the Coursewriter language so that an IBM 1440 CAI system could serve as the experimenter. CAI was chosen to administer the experiment because it would give a completely standardized set of instructions to each student and it would serve several students simultaneously.

When a student came to the experimental session a proctor conducted him to a CAI typewriter-terminal located in a booth. The proctor explained the use of the terminal, had the student type his name, and explained the use of the EOB signal which signified the end of a student response.

The following dialogue between computer and subject exemplifies the experimental procedure:

```
type sign on
sign on
type course name
concept 1
type your number
s1006
your name is      G. Biggs
Please type your group number and enter an EOB
signal.
3
Thank you.
Now open the book on the desk and read page one.
When you have finished, enter an EOB signal.
Just as all animals may be classified as vertebrate
or invertebrate so, in a like manner, all cards
indicative of (BLACK CIRCULAR FIGURES) may be
grouped together to form the concept (BLACK
CIRCULAR FIGURES).  Cards 17 and 40 are two of
several cards which belong to this concept.  Now,
you pick another card which belongs to the (BLACK
CIRCULAR FIGURES) concept, and type the number
of the card you pick.
REMINDER!  Don't forget to enter an EOB signal
after you type the card number.
4
The card whose number you typed is NOT an example
of the concept (BLACK CIRCULAR FIGURES).
Pick a different number.
11
Very good!
Pick another number.
18
Very good!
Pick another number.
25
Very good!
Now pick a card which does NOT belong to the (BLACK
CIRCULAR FIGURES) concept, and type the number of
the card you pick.
1
O.K., pick one more card which does not belong to
the concept.
2
That's fine.
Now read pages two, three, and four in the book and
```

```
enter an EOB signal when you have finished.
To make sure you know how to enter your hypothesis
correctly, type the relevant letter combination
for (BLACK CIRCULAR FIGURES).
jk
Very good!
We are now ready to begin the first task.  If
you have any questions now or later, you may go
back and read the instructions I have just given
you.  Before we begin please look at the clock
and type the time to the nearest minute as
accurately as possible.
1223
Thank you.
The focus card is number 25.  Select a card and
type the number of the card you select.
1
No. Pick another number or type hyp.
35
Yes.  Pick another number or type hyp.
44
Yes.  Pick another number or type hyp.
hyp
You are now ready to enter an hypothesis.  Type
the letter combination that corresponds to the
hypothesis you have chosen.
hk
Your hypothesis is incorrect.
The focus card is number 25.  Select a card and
type the number of the card you select.
18
Yes.  Pick another number of type hyp.
hyp
```

The array board was mounted on a wooden easel beside the typewriter terminal. A 5 x 9 inch hypotheses card was on a table at the side of the terminal. This card is reproduced below:

<u>ATTRIBUTE</u>	<u>VALUE</u>
Number of Borders.	.one = a
	two = b
Continuity of Borders.	.solid = c
	broken = d
Number of Figures.	.one = e
	two = f
Size of Figures.	.large = g
	small = h

Color of Figures. . . . . .white = i
black = j
Shape of Figures. . . . . circle = k
ellipse = l

To enter an hypothesis, type the letter combination that describes the relevant attribute values and then enter an EOB signal.

For Example: The relevant attribute values that describe the hypothesis for the concept (BLACK CIRCULAR FIGURES) are j for black and k for circular. This hypothesis would be entered in the following form and followed by an EOB signal.

The booklet of instructions which the subjects read when told to do so by the computer is reproduced below. These instructions follow closely those used by Ramsey (38).

INSTRUCTIONS FOR CONCEPT FORMATION TASKS

This experiment is concerned with how people attain concepts. You are asked to work several exercises in concept attainment. Your performance on these tasks is not related to your course grade.

On the easel before you is a display of 64 cards. Each card contains six attributes. The attributes are number of borders, type of borders, number of figures, size of figures, color of figures, and shape of figures. Each card has either one or two borders, broken or solid borders, one or two figures, large or small figures, white or black figures, and circular or elliptical figures.

Every card on the board is different from every other card in at least one of the six attributes just described. However, there are a number of ways certain cards may be grouped so that all cards in a specific group possess one or more of the same attributes. To illustrate from the animal kingdom, we know that all animals can be classified into two distinct groups--vertebrate or invertebrate. Thus, every animal either does or does not have a backbone and may be put into either group--vertebrate or invertebrate. When you have reached this point in your reading enter an EOB signal.

REMEMBER!! After each response, you must enter an EOB signal. (End of Page 1 of Directions)

You have just seen an example of a concept that has two relevant attributes, the two relevant attributes being color and shape of figures. The correct way to identify the concept is shown at the bottom of the attribute card on the desk to your right (look at the card carefully before you continue reading). For color of figures "j" was chosen and for shape of figures "k" was chosen because the concept is (BLACK CIRCULAR FIGURES). None of the values for the other four attributes was chosen since they were not part of the concept. In subsequent problems when you wish to offer an hypothesis about a concept you will type "hyp" and then enter an EOB signal. You should then type the identifying letters of the attributes that you think define the concept. Refer to the attribute card to find the identifying letters of correct values of the attributes.

All of the concepts you are to obtain have only two relevant attributes, therefore, you will never type more than two letters when you test an hypothesis.

In this experiment your job is to attain the concept that I have selected. At the beginning of each task I shall type the number of one card which belongs to a concept that I have selected and which you are to attain. This card we will call the focus card. You are to select cards which you want to test as belonging to the same concept as the focus card. Select each of your successive cards by typing the number of the card you chose and enter an EOB signal. After each selection I shall type "yes" or "no" depending upon whether or not the card you selected belongs to the concept. As you find which cards do and do not belong to the concept you can ascertain the concept.

Whenever you think you know the concept type "hyp" and I will let you enter your hypothesis in

> the manner described above. If your hypothesis is correct, the task is completed and we will go on to the next one. If not, I'll say "not correct" and you will continue selecting cards until you again think you know the concept. You may present as many hypotheses as you like. The job is to attain the concept as quickly as possible. When you have reached this point press the EOB signal.

Because of difficulties in scheduling computer time and because a number of students who were unfamiliar with CAI could not attain the concepts, approximately half the students (N = 29) were run with a human experimenter. He followed the computer procedure as closely as possible, but he did vary his explanations when students appeared not to comprehend the task.

The four concepts were arranged into four sequences and approximately one-fourth of the students completed each sequence. The N's for the four sequences were 13, 14, 12 and 12.

Analysis

Because there were only sixteen students in the pilot study, no formal statistical analysis of their data was made. The subgroup means for total card choices are presented in the next section of this paper.

In the major study a preliminary analysis of the means of the CAI experimenter versus the human experimenter groups and an analysis of the means of the different sequence groups indicated no significant differences in either case. These independent variables were ignored in subsequent analyses. For each pair of tests (verbal or figural closure and verbal or figural classification) the students in each treatment group were divided into high-high, high-low, low-high, and low-low subgroups. The general linear hypothesis method of statistical analysis was used because these subgroups did not contain the same numbers of students. The dependent variable was number of card choices to the correct hypothesis. Only the last two concept attainment problems and their total were considered in the analyses because intercorrelation matrices for the four problems computed separately for each treatment group indicated that performance on only the last two problems was sufficiently stable to give reliable results. Number of hypotheses was also considered as a dependent variable but the intercorrelation matrices for the four problems for this variable indicate that it was not a reliable performance measure. These intercorrelations matrices and the general linear hypothesis analysis are presented in the next section.

C. Results

The means for the subgroups of the pilot study are presented in Table 46. Support for the aptitude treatment interaction (ATI) hypothesis is best seen in the means of treatment by Hidden Patterns Test groups averaged over Four Letter Words. For the figural treatment, persons who were high in figural closure or cognition of figural units did better than persons who were low in this ability. For the verbal treatment, the reverse pattern is true.

TABLE 46

Subgroup Total Card Choice Means for Figural and Verbal Concept Attainment by High and Low Classification on Verbal and Figural Closure Tests

N = 16

	Treatment					
Tests	Figural			Verbal		
	Hidden Patterns Test			Hidden Patterns Test		
Four Letter Words	High	Low	Total	High	Low	Total
High	22.5	28.0	25.31	29.0	24.5	24.81
Low	25.5	35.0	30.31	36.0	21.0	28.51
Total	24.0	36.5		37.5	22.81	

Support for the ATI hypothesis is not presented in the pattern of means in the treatment by *Four Letter Words* groups averaged over *Hidden Patterns Test*. Persons high in verbal closure or ability to cognize symbolic units made better scores than persons low in this ability regardless of treatment.

The results of the statistical analyses of the data from the major study appear in Tables 47 through 50.

TABLE 47

Intercorrelations of Card Choice Scores for Four Concept Attainment Problems for Verbal and Figural Treatments

	Verbal				Figural		
	Problem				Problem		
	2	3	4		2	3	4
1	.09	.25	-.07	1	.21	-.01	-.16
2		-.01	.23	2		.26	.01
3			.51	3			.56

TABLE 48

Intercorrelations of Number of Hypotheses Scores For Four Concept Attainment Problems for Verbal and Figural Treatments

	Verbal				Figural		
	Problem				Problem		
	2	3	4		2	3	4
1	.12	.16	.02	1	-.30	-.13	-.19
2		.25	.57	2		.49	.23
3			.01	3			-.20

TABLE 49

F-Ratios for General Linear Hypothesis Analysis of Verbal-Figural Treatment by High-Low Figure Classification by High-Low Letter Sets Design

Source of Variation	Degrees of Freedom	Concept Three	Concept Four	Total Concept
Figure Classification (F)	1	6.84*	1.01	3.81
Letter Sets Test (L)	1	.05	.01	.03
Treatment (T)	1	.18	.59	.50
F x L	1	.15	.03	.09
F x T	1	.01	.01	.02
L x T	1	.23	.09	.00
F x L x T	1	.03	.60	.34
Error	45			

*significant beyond .05 level

TABLE 50

F-Ratios for General Linear Hypothesis Analysis of Verbal-Figural Treatment by High-Low Hidden Patterns by High-Low Four Letter Words Design

Source of Variation	Degrees of Freedom	Concept Three	Concept Four	Total Concept
Hidden Patterns Test (H)	1	2.71	1.52	2.65
Four Letter Words (W)	1	.18	1.95	1.20
Treatment (T)	1	.09	.39	.05
H x W	1	.51	2.49	1.84
H x T	1	.24	.40	.02
W x T	1	.82	.00	.21
H x W x T	1	.02	.02	.00
Error	45			

D. Discussion

The statistical analyses give no support to the ATI hypothesis. Furthermore, only Figure Classification showed a relationship to concept attainment. A number of reasons for failure to support the ATI hypothesis were apparent. First, and probably of most importance, was the lack of consistent performance by the students on the concept attainment tasks. The low interconcept correlations indicated that chance was an important determiner of success on the tasks. Second, the ability measures were not the precise ones indicated by the task analysis. Third, the verbal tasks may not have been semantic in nature.*

It is suggested that future studies of ATI using this kind of concept attainment task should require the students to attain six or eight concepts in order to assure obtaining reliable measures of performance. A re-analysis of the task should be made to determine if more appropriate aptitude measures could be chosen. In addition, further modifications of the tasks might lead to a sharpening of the differences between the treatments.

The use of CAI as a substitute for the human experimenter seems to be feasible. But it is recommended that if these procedures are to be used in the future, additional time be spent to train subjects in the use of the CAI student terminal.

E. Summary

Two studies were conducted in which performance on verbal and figural concept attainment tasks were related to aptitude variables which were believed to be relevant to the concept tasks. The concept tasks were presented by either computer assisted instruction facilities or by a human experimenter.

As a result of a task analysis based on the SI model, it was hypothesized that figural concept learning would be related to two figural ability factors, CFU and CFC, performance on the verbal concept task would be related to a semantic factor, CMC and a symbolic factor, CSU.

---

*Dr. Ralph Hoepfner. Personal communication. Assistant director, Aptitudes Research Project, University of Southern California, 1966.

A pilot study was conducted to test out the experimental procedures. No statistical analysis was made of these data but the results tended to support the ATI theory. In the major investigation, 24 students were presented the verbal concept tasks while 27 students were presented with the figural concept task. Approximately half of each of these groups were administered the tasks either via computer or human experimenter.

Four concepts, presented in four different sequences, were attained by each student. Dependent variables were the number of instances selected and the number of hypotheses offered. No differences were found between the computer versus human experimenter groups or between different sequences of presenting the concepts. The statistical analyses which tested the interaction of treatments and ability patterns gave no support for the ATI theory. Failure to support the theory might have been due to unreliability of the criterion measures, unfamiliarity of students with CAI equipment, or to an inadequate task analysis which resulted in the choice of inappropriate ability measures.

VIII. VOCABULARY LEARNING

In the concept formation study reported in the preceding section the following hypotheses, based on the aptitude treatment interaction (ATI) theory, were tested: (a) no difference between treatments when heterogeneous subject groups are compared, and (b) significant differences between the treatments when subjects are classified according to aptitude patterns believed to be relevant to the tasks. No support for the theory was obtained.

The general purpose of the studies reported in this section was to test further the second prediction mentioned above. These studies are based on vocabulary learning tasks which are probably more meaningful to students than are concept formation tasks. Also, the task analysis procedures for determining relevant abilities were improved.

The first study deals with the identification of abilities which appear to be of importance in learning vocabulary words by each of two methods. The second study is a pilot investigation of the interaction of these abilities and three methods of teaching vocabulary (of which two were developed in study one). The third study is a large-scale replication of the second study and it includes revised learning materials. The fourth study is similar to the preceding two except that the interaction of ability and method is investigated over a prolonged period of training.

Cronbach and Gleser (12) state that the comparison of regression coefficients is more appropriate in exploring ATI than is the comparison of correlation coefficient for the following reasons: First differences in correlations may result from differences in variability only and second crossover points for the regression lines may be computed to determine whether they cross within the range of scores of the independent variable. If they do not, then one treatment is always better than the other. In the studies reported in this section, correlations rather than regression coefficients were used because of the exploratory nature of the studies and the large number of comparisons to be made. In addition, the criterion measures favored one of the treatments and thus reduced the possibility that the crossover point of the regression lines would occur within the range of ability scores when that treatment was involved.

A. Identification of Relevant Abilities for Two Methods of Learning Vocabulary Words

Johnson and Stratton (31) demonstrated that teaching the meanings of words can be done equally well by four methods. These methods are presenting definitions directly, having students derive word meanings from context, giving a series of short sentences that relate to the words to be learned (classification method), and utilizing synonyms. They also used a mixed method which was superior to the others.

The Johnson and Stratton study served as a point of departure in designing the study described below. The purpose of this study was to identify abilities which seemed to be necessary for learning vocabulary by two different methods of instruction.

Method

Materials. Two kinds of verbal concept materials were devised. These corresponded to the definition and classification methods used by Johnson and Stratton. In the definition material, the meanings of the words to be learned were simply stated in expository form. The student's task was to read the definitions and to rewrite them in his words.

In the classification material, a series of sentences relating to the words to be learned were presented in an inductive manner. The subject responded to each sentence by writing next to it a word, selected from several, that he thought best fitted the meaning. When he did this, the correct answer was identified and eventually, after presenting the series of sentences, the correct answer (concept) was learned.

These words were to be learned; limpid, duress, lissome, salubrious, nascent and paroxysm. They were regarded as appropriate for the study because they were defined correctly in a multiple choice vocabulary test (V-4 of the Kit of Reference Tests) by less than 20 percent (chance-level) of a sample of 80 undergraduate students who were enrolled in an educational psychology course.

In the definitions material, each word was used in a rather specific way in a sentence and it was followed by a more general definition in a second sentence. Each student then wrote the definition in his words.

In the classification material, six short sentences for each word were constructed. Each set of six sentences (1 for each word) was typed on a separate page. On each succeeding page, the correct answers for the sentences of the previous page were presented and a new set of sentences was presented. Each page contained a different order of sentence (meaning) presentation.

Subjects. Registrants in a graduate-level class were used as subjects. Six of them received the definition material and another six received the classification material.

Procedure. After the students completed the learning tasks, they were presented a list of 24 words that were synonyms of the six words that were to be learned. The students were directed to write the appropriate word, of those that were learned, after each synonym. There were four synonyms for each word.

After the criterion test was completed each student was given 16 slips of paper. Each contained the definition of an ability from Guilford's Structure of Intellect (SI) model and an example in the form of a typical test item which is used to assess it. Most of the definitions were taken from Allen, Guilford and

Merrifield (1). The 16 abilities were selected so that all were from the semantic dimension of the SI model, a test was available for each, and the ability was verbally defined. Sixteen abilities met these specifications and they appear in Table 51.

TABLE 51

Abilities Used in Pilot Study I According to Guilford's SI Model

Model Label		Verbal Description
CMU	--	Verbal comprehension
CMC	--	Conceptual classification
CMR	--	Semantic relations
CMS	--	General reasoning
CMT	--	Penetration
CMI	--	Conceptual foresight
MMU	--	Memory for ideas
NMU	--	Concept naming
NMC	--	Convergent production of semantic classes
NMR	--	Semantic correlates
NMS	--	Semantic ordering
NMT	--	Semantic redefinition
NMI	--	Deduction
EMU	--	Evaluation of Semantic Units
EMC	--	Evaluation of Semantic classes
EMR	--	Evaluation of Semantic relations

Results

Each student ranked the abilities according to the extent he thought each would be important in learning the meanings of the words. A rank of 16 was assigned to the most important ability. The average rank totals for the abilities for each of the two sets of materials appear in Table 52. For comparison

TABLE 52

Average Rank Totals for the 16 Abilities for the Two Types of Materials

	Classification (N=5)	Definitions (N=6)
CMU	11.0	15.2
CMC	11.0	10.7
CMR	11.2	11.3
CMS	4.8	5.8
CMT	7.4	5.8
CMI	2.0	4.3
MMU	7.8	10.7
NMU	10.6	11.3
NMC	13.6	12.0
NMR	8.2	9.3
NMS	5.6	3.3
NMT	8.2	4.3
NMI	4.8	8.0
EMU	12.2	9.5
EMC	9.4	7.7
EMR	8.2	8.2

purposes, the top-ranked 7 abilities, as viewed by each of the groups, appear in Table 53.

TABLE 53 *

Seven Top Ranking Abilities as Viewed by Two Groups of Judges

Classification		Definitions	
Ability	Rank	Ability	Rank
NMC	1.0	CMU	1.0
EMU	2.0	NMC	2.0
CMR	3.0	CMR	3.5
CMU	4.5	NMU	3.5
CMC	4.5	MMU	5.5
NMU	6.0	CMC	5.5
EMC	7.0	EMU	7.0

*The seven top abilities according to each group were reranked according to the average rank total given in Table 53.

Discussion

The two groups of students generally perceived the same abilities as being most important for learning the material. Only EMC (Evaluation of semantic relations) and MMU (Memory for ideas) were not common to the top rankings of the groups. However, the groups differed with regard to how they ordered the abilities which were perceived to be most important. The rank order correlation of the six commonly selected abilities was -.34 which suggests that there is a somewhat different ability pattern needed to learn each of the two tasks.

3

In general, both groups identified primarily verbal meaning abilities (verbal, comprehension, conceptual classification, semantic relations, concept naming, convergent production of semantic classes and evaluation of semantic units) as being important but they placed them in different rank orders.

The relatively high agreement within groups coupled with the different orders of ranking between groups seemed to warrant the use of the ability measures in future empirical studies of ATI.

B. Pilot Investigation of Three Methods of Learning Vocabulary Words

In the preceding study, eight abilities were selected as being relevant to learning vocabulary words by two different methods. Six of the abilities were selected as important for both methods but their rank orders differed. One ability which was specific for each method was identified.

The general purpose of the second study was to determine empirically the relationships between these eight abilities and achievement of vocabulary material under the two methods of the first study plus one other method devised by Johnson and Stratton (31). The specific purpose was to try-out the tests and materials before administering them to a large group of subjects.

Method

Subjects. Twenty-six students from grades nine through twelve were used.

Materials. Three sets of learning materials were used in this study; definition, classification, and synonyms.

The definition material consisted of verbal definitions for 4 different and difficult words--paroxysm, salubrious, altercation and alacrity. Each word was defined somewhat specifically in one sentence and it was followed by a more general statement of definition in a second sentence. An example of this procedure is given below.

> A person undergoing a visible and violent reaction from either physical or emotional causes can be said to be having a paroxysm. Thus a paroxysm is any unusually explosive excitation.

A space followed the definition and the student was to write in it the definition in his words.

In the classification material, the meanings of 4 words were taught inductively. On each page (six pages in all) there were five short sentences each of which gave a clue to the definition of a particular word. The words to be learned appeared at the top of the page and each was to be written in the space next to the sentence that pertained to it. Although five sentences appeared on each page, only four of them dealt with the word to be learned. After the student matched the words to the sentences, he folded back the sheet to determine if his choices were correct or not. This procedure was used on all pages. This approach provided a repeated trial and error matching of words to sentences, and the assumption was that correct definition would eventually be induced. The four words used in the classification material were: duress, limpid, lissome and nascent. A sample page is given below.

List 1: duress, limpid, lissome, nascent

List 2:		(On back of page)
1.	________ A tulip bulb shows just a tip of green above the earth.	nascent
2.	________ A tall tale told by a fisherman.	none
3.	________ A hula dancer performs with grace and skill.	lissome
4.	________ Direct rays from the sun light up a tiny glade encircled by deep forest	limpid

5. ________ A wild bird trys to escape through the bars of his cage. duress

In the synonym material the same basic procedure was used as with the classification material except that single word synonyms were utilized instead of short sentences. The 4 words learned by this method were: antipodal, ignominious, stripling and succinct. One of the six pages which were used in the study is reproduced below.

List 1: antipodal, ignominious, stripling, succinct

List 2: (On back of page)

1. ________ lad stripling

2. ________ dishonorable ignominious

3. ________ short succinct

4. ________ heal none

5. ________ opposite antipodal

Nine of the 12 words which were included in the pilot study had Thorndike-Lorge (42) counts of less than 1 in a million. Three of the words--antipodal, lissome and salubrious--were not in the Thorndike-Lorge list.

Tests. Two criterion measures were constructed for the words of each method. The first test, synonym production, consisted of the relevant words and spaces in which the subjects were instructed to write as many synonyms as they could within a five minute time period. This test was also used as a pretest. The second test, synonym matching, consisted of the list of words to be learned followed by a list of synonyms in random order. The subjects were instructed to write the number of the appropriate word by each synonym. Five minutes was the time limit for this test. Complete sets of learning materials and criterion tests are shown in Appendix C.

A battery of nine ability measures was used. Controlled associations, a measure of divergent production of semantic relations (DMR), was included with tests of the eight abilities previously identified, because it was felt that it might show differential correlations over the three methods with the synonym production test. The 9 ability measures and their respective SI factors are shown in Table 54.

TABLE 54

The Nine Ability Measures and Their Respective Factors Used in Pilot Study II

Tests	Factors
1. Word classification	Conceptual classification (CMC)
2. Verbal analogies	Semantic relations (CMR)
3. Wide range vocabulary (V-3)	Verbal comprehension (CMU)
4. Controlled associations	Associational fluency (DMR)
5. Class name selection	Evaluation of semantic classes (EMC)
6. Double descriptions	Evaluation of semantic units (EMU)
7. Memory for word meanings	Memory for ideas (MMU)
8. Word grouping	Convergent production of semantic classes (NMC)
9. Word-group naming	Concept naming (NMU)

Procedure. Two testing days were used. The ability measures and the synonym production test were administered on the first day. The learning materials and criterion measures were administered on the second day.

The 9 pretest ability measures were arranged in two test booklets. The booklet which was administered first contained the Verbal Analogies, Double Descriptions, Class Name Selection, Word Classification, Memory for Word Meanings and Wide Range Vocabulary tests all of which yielded machine scorable responses. After a short break, the second booklet--consisting of the Controlled Associations, Word Group Naming and Word Grouping tests--and the synonym production tests were administered.

On the second day the three sets of learning materials were administered. Each student worked through all sets. Three presentation orders were used. All students were asked to indicate the time it took them to work through each set of materials. Spaces on which to write beginning and ending times were placed on each set of materials.

The criterion tests were administered after all students completed the three sets of materials.

Results

The correlations between ability measures, two criterion tests (synonyms production and synonym matching) and time spent on all three types of learning material were of major concern.

Table 55 contains correlations between the ability tests, time measures, and the synonym matching test. The critical aspect is the differential relationships between the ability measures and performance on the criterion test for the three treatment groups. The evaluation tests (EMU and EMC) were more highly related to performance based on the synonym and classification materials than on the definition materials.

TABLE 55

Correlations of Ability and Time Scores with Synonym Matching Performance

Ability Tests	Type of Material Definition	 Classification	 Synonym
Verbal Analogies (CMR)	.62**	.62**	.55**
Double Description (EMU)	.60**	.65**	.77**
Class Naming (EMC)	.31	.51**	.68**
Word Classification (CMC)	.59**	.77**	.71**
Memory (MMU)	.38	.37	.61**
Vocabulary (CMU)	.73**	.57**	.52**
Controlled Associations (DMR)	.53**	.64**	.70**
Word Group Naming (NMU)	.62**	.79**	.73**
Word Grouping (NMC)	.48*	.47*	.61**
Time on Definitions	.25		
Time on Classification		-.41*	
Time on Synonyms			-.81**

*Significant at the .05 level

**Significant at the .01 level

The memory test (MMU) is more highly related to performance based on the synonym materials. Vocabulary (CMU) best predicts performance based on the definition material. The convergent production tests (NMU and NMC) are more highly related to performance based on the classification and synonym materials. The time measures correlate differently with the criterion measure. The highest negative correlation is with performance based on the synonyms material.

Table 56 contains the correlations between the ability and time measures and performance on the synonym production criterion test. The obtained patterns of interaction are similar to those obtained for the synonym matching criterion. The major difference is the relationship between the vocabulary (CMU) ability measures and performance. In this case, vocabulary is not a differential predictor but is significantly related to achievement by all methods.

Discussion

The results of the study appear to support the ATI prediction that different methods of instruction involve different patterns of abilities. It must be noted, however, that the rank order of correlations for the definitions and classifications materials do not correspond particularly well with the rank orders assigned by the graduate student judges. The small sample size of this study makes detailed interpretation of the results unwarranted. These results do indicate, however, that the materials and procedures appear to be suitable for high school and college freshman students.

TABLE 56

Correlations of Ability and Time Scores with Synonym Production Performance

	Type of Material		
Ability Tests	Definition	Classi-fication	Synonym
Verbal Analogies (CMR)	.55**	.45*	.47
Double Description (EMU)	.37	.50**	.67**
Class Naming (EMC)	.16	.40*	.62**
Word Classification (CMC)	.54**	.70**	.67**
Memory (MMU)	.46*	.49**	.51**
Vocabulary (CMU)	.60**	.55**	.55**
Controlled Associations (DMR)	.57**	.62**	.61**
Word Group Naming (NMU)	.59**	.81**	.63**
Word Grouping (NMC)	.59**	.46*	.60**
Time on Definitions	.22		
Time on Classification		-.41*	
Time on Synonyms			-.66**

*Significant at the .05 level

**Significant at the .01 level

C. An Investigation of Three Methods of Learning Vocabulary Words

The purpose of this study was to determine whether differential patterns of correlations exist between a set of nine ability measures and achievement under three methods for learning vocabulary words. The nine ability measures, the instructional materials and the criterion measures were the same as those employed in the previous study.

Method

Subjects. The subjects were 164 college freshmen. The sample included approximately equal numbers of males and females.

Materials. The materials and the procedures for administering them were identical to those of the previous study. Complete sets of materials and criterion test are shown in Appendix C.

Results

Descriptive results for the ability measures are presented first. Table 57 gives the means and standard deviations of the 164 students on the nine ability measures and Table 58 presents the intercorrelations of the nine variables.

Because it was consistently found that average pretest performance was near zero, only posttest scores were considered in the analysis. Table 59 contains the means and standard deviations of the time and criterion measures for the group.

TABLE 57

Means and Standard Deviations of the Nine Ability Tests

(N=164)

Test	M	SD
Verbal Analogies	16.54	4.20
Double Descriptions	32.84	5.31
Class Naming	13.71	1.99
Word Classification	12.18	2.60
Memory for Word Meanings	25.80	8.75
Wide Range Vocabulary (V-3)	11.77	3.93
Controlled Associations	23.98	6.70
Word Group Naming	18.26	3.81
Word Grouping	28.07	7.11

TABLE 58

Intercorrelations of the Nine Ability Measures*

	1	2	3	4	5	6	7	8	9
Verbal Analogies	1.00	.29	.18	.28	.20	.35	.26	.28	.27
Double Descriptions		1.00	.40	.38	.25	.27	.25	.28	.24
Class Naming			1.00	.30	.26	.28	.14	.22	.06
Word Classification				1.00	.29	.43	.14	.23	.28
Memory for Word Meanings					1.00	.19	.24	.09	.28
Wide Range Vocabulary (V-3)						1.00	.39	.35	.27
Controlled Associations							1.00	.38	.25
Word Group Naming								1.00	.18
Word Grouping									1.00

*Correlations $\geqq$ .16 significant at .05 level.

TABLE 59

Means and Standard Deviations of the Criterion Tests and Average Times to Complete the Learning Materials

N=164

		Definition	Classification	Synonym	Total
Synonym Matching	M	10.70	11.31	13.77	35.78
	SD	4.18	3.74	2.89	9.24
Synonym Production	M	9.39	7.71	17.24	34.34
	SD	6.41	5.55	6.78	14.56
Time	M	4.20	7.59	6.20	17.98
	SD	1.64	2.12	1.97	3.60

The differential correlations between each ability test and the two criterion tests are of greatest importance. Table 60 contains the correlations between the time and ability measures with synonym matching performance based on each type of instructional materials.

All ability tests, except the Memory for Word Meanings Test, are significantly related at the .01 level to total synonym matching performance. The Wide Range Vocabulary Test is the best overall predictor. Of interest also is the negative correlation between total time and total achievement. Although almost all tests show differential predictions with achievement based on the different types of material, only some of the more interesting comparisons are mentioned here.

TABLE 60

Correlations of Ability and Time Scores with Synonym Matching Performance

N=164

Ability Tests	Type of Material Definition	 Classification	 Synonym	 Total
Verbal Analogies (CMR)	.17*	.26**	.28**	.27**
Double Description (EMU)	.14	.23**	.23**	.24**
Class Naming (EMC)	.21**	.29**	.20**	.28**
Word Classification (CMC)	.19*	.18*	.13	.21**
Memory (MMU)	.00	.04	-.09	.01
Vocabulary V-3 (CMU)	.47**	.57**	.45**	.58**
Controlled Associations (DMR)	.28**	.38**	.24**	.35**
Word Group Naming (NMU)	.23**	.29**	.18*	.27**
Word Grouping (NMC)	.16*	.19*	.17*	.21**
Time on Definitions	.00			
Time on Classification		.01		
Time on Synonyms			-.37**	
Total Time				-.21**

*Significant at the .05 level

**Significant at the .01 level

Verbal Analogies, although related to performance on all three sets of material, is more highly related to the classification and synonym methods than to definitions. This relationship is reasonable because in the classification and synonym material students are likely to have to think in terms of how sentences or synonyms are alike. This kind of thought could very well involve the cognition of semantic relations. The definitions materials, on the other hand, would appear to depend to a lesser extent on this ability. Double Descriptions, an evaluation test, predicts performance based on classification and synonym materials but not performance based on definition materials. This is reasonable because in the definition method the definitions are stated and examples are given--therefore, they call for less of a cognitive evaluation process. Word Classification, a test of ability to recognize common properties of words, ideas, etc., does not relate to the criterion in the expected manner. It was believed to be more highly related to classification and synonym materials than definition materials. Memory for Word Meanings is not related to performance based on any materials. The Wide Range Vocabulary Test is highly related to performance based on all types of material. These associations were expected because all the tasks are vocabulary learning. The Controlled Associations and Word Group Naming tests, being measures of production factors, are more highly related to performances based on definition and classification materials than the synonym method. This was expected because few or no synonyms are presented in either of these materials. However, the Word Grouping test did not conform to these expectations although it is also a measure of a production factor.

Table 61 presents the correlations between the time and ability measures with posttest synonym production achievement on each of the three types of vocabulary materials. Again all the ability tests except Memory for Word Meanings are significantly related at .01 level to total criterion performance with Wide Range Vocabulary being the best overall predictor.

TABLE 61

Correlations of Ability and Time Scores with Synonym Production Performance

N=164

Ability Tests	Type of Material Definition	 Classification	 Synonym	 Total
Verbal Analogies (CMR)	.20**	.35**	.20**	.31**
Double Description (EMU)	.17*	.17*	.23**	.24**
Class Naming (EMC)	.14	.17*	.20**	.21**
Word Classification (CMC)	.13	.28**	.15	.23**
Memory (MMU)	.10	.16*	.00	.10
Vocabulary V-3 (CMU)	.47**	.43**	.37**	.54**
Controlled Associations (DMR)	.43**	.28**	.30**	.43**
Word Group Naming (NMU)	.19*	.23**	.22**	.27**
Word Grouping (NMC)	.30**	.27**	.16*	.30**
Time on Definitions	.00			
Time on Classification		-.16*		
Time on Synonyms			-.29**	
Total Time				-.27**

*Significant at the .05 level

**Significant at the .01 level

The same general pattern of correlations with the synonym production criterion test was obtained as with the synonym matching test except for the following differences. Class Naming is significantly related to performance on the classification and synonym materials but not to performance on the definition material. Controlled Associations is more highly related to the definition material than was found with the synonym matching test. This again would be predicated, especially when the criterion test requires the students to produce synonyms which were not available in the learning material. Word Grouping, although significantly related to all materials on the matching test, is more highly related when the criterion test is production in nature.

Discussion

The results of the study generally support the ATI theory but they are not consistent with the judgements of graduate students nor are they always consistent with expectations based on logical or theoretical considerations. For example, one might hypothesize that achievement of the synonym materials would be more dependent on short term memory than achievement of either definitions or classification materials regardless of the criterion used. The findings of the pilot study were consistent with this conjecture but the major study failed to corroborate it. Instead, achievement of classifications material, when the criterion test was synonym production, was the only measure related to memory. It is possible that the memory test, of memory for semantic units, was not as appropriate as a test of memory for semantic classes might have been.

Each of the nine ability measures appears to be involved in one or more aptitude treatment interactions although formal tests of significance of differences between correlations were not made. Future studies of ATI effects with these kinds of materials should use criterion measures that do not favor one group and should employ regression analysis procedures.

D. Two Methods of Vocabulary Learning at Different Stages of Practice

The purpose of this study was to determine whether different patterns of correlations of ability tests and achievement under two methods of learning vocabulary words would be modified by practice with the instructional method. Fleishman and his associates (19, 20, 21) and Kohfield (33) have shown differential relationships between ability measures and psychomoter performance at different stages of practice. Recently Bunderson (8) and Dunham, Guilford and Hoepfner (14) have also shown differential relationships between SI factors and performance on non-psychomoter laboratory tasks at different stages of practice. In all of these studies the task that was practiced did not change. In the present study different words were learned at each practice stage with only the method of learning being constant. In this respect the present study resembles the work of Harlow (30) on learning sets more than the studies previously cited. It should be pointed out that the research which is reported here is a pilot investigation and, like most try-outs, contains many flaws. A larger more comprehensive study is planned for the near future but will not be done in time to report here.

Method

Subjects. Subjects for this study were two classes of 11th grade English students. An average ability-level group was used for the synonym material and an honors section was used for the definitions material. Conditions over which the investigators had no control prevented them from forming groups randomly.

Materials. Two sets of instructional materials on vocabulary were developed, definitions and synonyms. Each is discussed below. Both sets have in common forty vocabulary words which were selected from the Thorndike-Lorge (42) handbook.

Because difficult words with which none or few students would have familiarity were desired, every word was taken from the lowest frequency count category. However, no word was selected unless there were six synonyms for it.

The synonym method materials were learned by inductive presentation. Each set of these materials contained four words to be learned. The four words were at the top of each page and were followed by a list of five other words. Four of these words were synonyms to the key learning words and one word was not an appropriate synonym. A blank space was placed beside each synonym and the students wrote in it what they thought to be the word which matched the synonym. The word "none" was written next to the extra word. The correct answers were on the back of the page. Each set of four vocabulary words contained six pages as described above--every page utilized a different set of synonyms for the four vocabulary words. Ten different sets were constructed.

In the definition material, the same 10 sets of vocabulary words were used. The definitions for each of the four words within a set were written out. Each word was defined in a specific way in the first sentence and it was followed by a more general statement of the definition in a second sentence. A blank space was provided following both sentences, in which the student was to re-write the definitions in his words. The criterion measure for each set of materials for both groups was a synonym production test. Complete sets of materials and a sample criterion test are shown in Appendix C.

The reference battery consisted of tests representing eight factors postulated by the SI model. Because the materials were essentially the same as for the other vocabulary studies, the battery of tests was similar. Some additional tests were included and some deleted on the recommendation of Dr. Ralph Hoepfner of the Aptitudes Research Project, University of Southern California. Table 62 contains the ability tests, the factors they are intended to measure, and the SI code label.

TABLE 62

Ability Tests, SI Factors and Code Labels Used in The Study

Ability Tests	Factors	Code Label
Wide Range Vocabulary	Cognition of Semantic Units	CMU
Word Classification	Cognition of Semantic Classes	CMC
Memory for Ideas	Memory of Semantic Units	MMU
Classified Information	Memory of Semantic Classes	MMC
Topics Test	Divergent Production of Semantic Units	DMU
Utility Test	Divergent Production of Semantic Classes	DMC
Word-Group Naming	Convergent Production of Semantic Units	NMU
Word Grouping	Convergent Production of Semantic Classes	NMC

Procedure. The procedure used in the present study was similar to that used in the studies previously reported in this section. The reference battery was given first followed by the learning materials. The major difference was that the learning materials extended over a longer period of time. There were ten sets of material (both definitions and synonyms) with four different words in each set, one set was given each day for ten days.

On the first day following the administration of the ability tests, the first set of vocabulary words was given. During the first few days students were given a pretest before receiving the learning materials but this was abandoned because very few of them knew the meanings of any of the words. Subsequently only the synonym production posttest was administered immediately after the materials were finished.

The times at which each student started and completed the set of materials were noted. When everyone had completed the materials, the posttest was given. Three minutes of working time were allowed for it.

Originally it was planned to give students who used the definition material a posttest requiring them to write definitions and the synonym students a posttest requiring them to write synonyms. However, because the procedure for scoring the defintiions resulted in very little variability, students using the definitions material were switched to the same synonym production criterion posttest as the synonym group used.

Because some students had excessive absences, the analysis is based on 25 students in the synonym group and 15 in the definition group. Because time and posttest measures were obtained for each student on each of 10 days, a total of twenty scores were available. In several cases, if one data entry was not available (e.g. - a student failed to record his final time) the mean score for the other nine scores was entered. However, this procedure was used only if one or two data entries was missing. Students with more than two missing data points were eliminated from the analysis.

Results

For each student scores were obtained from eight ability tests, time to complete the learning materials, and posttest achievement in the synonym production test. Performance on each of the ability tests was correlated with both the synonym production test and time at each of the ten stages of practice.

Table 63 contains the means and standard deviations of the two classes on the eight ability tests. The results indicate that the honors class is consistently superior to the average class. Because

TABLE 63

Means and Standard Deviations for the Two English Classes on Eight Ability Tests Used in the Vocabulary Practice Stage Study

Tests	Average Class		Honors Class	
	M	SD	M	SD
Wide Range Vocabulary	9.48	3.07	14.80	3.31
Word Classification	9.64	2.35	11.67	2.68
Memory for Ideas	25.28	7.27	35.07	8.02
Classified Information	35.76	6.75	45.53	7.14
Topics Tests	21.20	5.19	26.00	7.12
Utility Test	11.86	6.57	15.13	6.38
Word-Group Naming	16.68	3.40	20.13	2.22
Word Grouping	26.04	4.74	31.80	7.05

comparisons of patterns of correlations rather than mean criterion performances are of major interest in this study the inequality of the two groups might not completely invalidate the study.

Table 64 presents the means and standard

TABLE 64

Means and Standard Deviations for the Synonym and Definition Groups on the Time and Synonym Production Measures for Ten Days

	Achievement				Time			
	Synonym		Definition		Synonym		Definition	
Day	M	SD	M	SD	M	SD	M	SD
1	12.52	3.97	*	*	8.48	1.68	3.67	1.30
2	16.48	4.84	*	*	7.40	1.90	4.00	1.67
3	15.36	4.38	9.80	2.76	6.84	1.74	4.20	.98
4	14.76	5.44	9.53	1.89	6.68	2.24	3.53	1.45
5	13.92	5.82	11.00	3.26	8.16	1.97	4.20	1.38
6	12.12	5.45	9.46	2.47	6.32	1.95	4.13	.88
7	10.76	4.55	9.60	2.06	8.40	1.72	3.80	.91
8	13.32	5.07	9.73	1.95	7.00	2.04	3.80	1.56
9	15.84	4.58	13.06	2.29	5.80	1.79	3.60	1.08
10	11.56	4.92	6.66	1.85	7.80	1.82	3.60	.88

deviations for both the synonym and definition groups on the time and synonym production posttest measures. Inspection of the average achievement scores reveals no tendency for the development of a learning set. That is, no consistent pattern of improvement throughout the ten days is evident. If anything, the synonym group tends to get worse, whereas the definition group seems to do equally well on all days, except the last two. These results might simply be a reflection of the differences between the materials on each day even though vocabulary words were randomly assigned. The time scores again show the same relative trends. The definition material, however, takes considerably less time to complete than the synonym material.

Table 65 presents the correlations between the memory (MMV and MMC) tests and posttest achievement.

TABLE 65

Correlations Between The Memory Tests and Posttest Achievement

Days	Synonym		Definition	
	MMU	MMC	MMU	MMC
1	.32	.29	a	a
2	.54**	.41*	a	a
3	.42*	.33	.06	.05
4	.51**	.41*	-.25	-.16
5	.32	.48*	.44	.37
6	-.27	.13	-.25	-.40
7	.33	.30	.18	.18
8	.48*	.33	-.20	-.09
9	.26	.31	.18	.51*
10	.31	.15	.16	.05

a Correlations ommitted because of different posttest
* Significant at the .05 level
** Significant at the .01 level

In the synonym groups memory is significantly related to achievement during the first stages of practice and then tends to become less important. In the definition group, the only significant correlation is found during the ninth practice day. There does seem to be an increasing trend in importance of the memory tests in predicting posttest performance in the definition group.

Table 66 presents the correlations between the

TABLE 66

Correlations Between the Cognition Tests and Posttest Achievement

Days	Synonym		Definition	
	CMU	CMC	CMU	CMC
1	.21	-.20	a	a
2	-.03	.22	a	a
3	.01	.18	.53*	.17
4	-.11	.02	.05	-.11
5	.26	-.17	.01	.00
6	.09	.00	-.18	.27
7	-.03	.01	.27	.04
8	.12	.00	.17	.28
9	.07	.15	-.02	.44
10	.16	.13	.12	.06

a Correlations omitted because of different criterion posttest

* Significant at the .05 level

cognition tests (CMV and CMC) and posttest achievement. In the synonym group no significant correlations were obtained at any stage of practice although there seemed to be a slight increasing trend. In the definition groups only one significant correlation was found, that being on their first day.

Table 67 presents the correlation between the divergent production tests (DMU and DMC) and posttest

TABLE 67

Correlations Between the Divergent Production Tests and Posttest Achievement

Days	Synonym		Definition	
	DMU	DMC	DMU	DMC
1	.29	-.14	a	a
2	.45*	-.06	a	a
3	.44*	.17	.01	-.45
4	.51**	.09	-.30	.04
5	.14	.27	.47	.22
6	-.34	-.19	-.24	-.05
7	.57**	.14	.11	-.18
8	.63**	.31	-.09	.09
9	.29	.04	.28	.46
10	.70**	.32	-.27	-.25

a Correlations omitted because of different criterion posttest

* Significant at the .05 level

** Significant at the .01 level

achievement. In the synonym groups, the divergent production tests have significant correlations throughout the practice stages. In addition the correlations tend to become larger through time. In the definition groups, no significant correlations were found.

Table 68 presents the correlations between the convergent production tests (NMU and NMC) and posttest

TABLE 68

Correlations Between the Convergent Production Tests and Posttest Achievement

Days	Synonym		Definition	
	NMU	NMC	NMU	NMC
1	.21	.27	a	a
2	.01	.34	a	a
3	.20	.51**	.23	-.37
4	-.07	.26	-.41	-.14
5	.16	.08	.19	-.19
6	.08	-.26	.21	.03
7	-.01	.02	-.16	.12
8	.25	.16	-.05	.21
9	-.08	.13	.01	-.30
10	.28	.32	-.51	.18

a Correlations omitted because of different criterion posttest

* Significant at the .05 level

** Significant at the .01 level

achievement. In the synonym group the only significant correlation obtained was at the third stage of practice. In the definition groups, the only significant correlation obtained was on the last day of practice. However, the correlation was negative.

Table 69 presents the correlations between the time and achievement measures at each stage of practice.

TABLE 69

Correlations Between Time and Posttest Achievement at Each Stage of Practice

Days	Synonym	Definition
1	-.34	a
2	-.27	a
3	-.32	.04
4	-.35	-.13
5	-.25	-.07
6	-.02	-.18
7	-.34	-.33
8	-.41*	-.09
9	-.11	-.26
10	.24	-.08

a Correlation omitted because of different criterion posttest

* Significant at the .05 level

In the synonym groups all correlations except one are negative and only one is significant. In the definition groups, only two of the correlations are positive but none of the correlations is significant.

Table 70 presents the correlations between the memory tests (MMU and MMC) and the time to complete

TABLE 70

Correlations Between the Memory Tests and Time to Complete the Learning Materials

Days	Synonym		Definition	
	MMU	MMC	MMU	MMC
1	.03	.06	.11	.39
2	-.30	-.11	-.25	-.11
3	-.40*	-.22	-.18	-.28
4	-.44*	-.26	-.32	-.50
5	-.16	-.25	-.01	.05
6	-.01	.23	-.05	-.01
7	-.44*	.04	-.05	-.15
8	-.38	-.02	.36	.14
9	-.60**	-.28	-.20	-.11
10	-.43*	.09	-.07	-.11

* Significant at the .05 level

** Significant at the .01 level

the learning materials. In the synonym groups the only significant correlations are negative and tend to increase over practice stages. In the definition groups none of the correlations is significant although the majority are negative.

Table 71 presents the correlations between the cognition tests (CMU and CMC) and the time to complete the learning materials. In the synonym groups, although there are no significant correlations, an apparent decrease from positive to negative correlations was

TABLE 71

Correlations Between the Cognition Tests and Time To Complete the Learning Materials

Days	Synonym		Definition	
	CMU	CMC	CMU	CMC
1	.27	.16	-.04	.29
2	-.09	-.08	.05	-.09
3	-.07	-.13	-.03	.28
4	-.04	-.01	.06	.35
5	-.27	.00	-.08	.27
6	-.25	.16	.19	.39
7	-.23	-.34	.45	.03
8	-.21	-.35	-.11	.08
9	-.38	-.17	.29	.39
10	.05	-.04	.18	.09

obtained. The correlations in the definition groups were mostly positive with no apparent trend being found. Also none of these correlations was significant.

Table 72 presents the correlations between the divergent production tests (DMU and DMC) and time to complete the learning materials. Most of the correlations in the synonym groups were negative, none was significant.

TABLE 72

Correlations Between the Divergent Production Tests and Time to Complete the Learning Material

Days	Synonym		Definition	
	DMU	DMC	DMU	DMC
1	.03	.08	-.24	.29
2	-.24	-.01	-.27	.21
3	.02	-.16	-.39	.16
4	-.38	-.31	-.26	.21
5	.17	.03	-.48	.25
6	-.20	-.39	-.28	.17
7	-.09	.00	-.41	-.27
8	-.26	-.21	-.40	-.06
9	-.21	-.19	-.41	.08
10	.16	.01	-.35	.06

Also no trend was apparent. In the definition groups, one of the tests (DMC) showed a trend from positive to negative correlations. The other test (DMU) in the definition groups showed a slight correlational trend from -.24 to approximately -.40. However, neither of the tests had any significant correlations with time.

Table 73 presents the correlations between the convergent production tests (NMU and NMC) and time to complete the learning materials. In the synonym groups the only significant correlation was negative, however a slight upward trend towards zero order correlations was obtained. No significant correlations were found in the definition group although again a slight upward trend from negative towards zero order was found.

TABLE 73

Correlations Between the Convergent Production Tests and Time to Complete the Learning Materials

Days	Synonym		Definition	
	NMU	NMC	NMU	NMC
1	-.01	-.05	-.03	-.39
2	.19	-.12	-.47	.30
3	-.21	-.01	.11	.18
4	-.54**	-.21	.29	.17
5	.11	-.11	-.31	-.07
6	-.03	-.02	.09	-.26
7	-.01	.07	-.05	.12
8	.07	.05	-.26	.02
9	.02	-.27	.22	-.25
10	.14	-.02	-.11	.14

** Significant at the .01 level

Discussion

Because of the small, dissimilar samples used in the study the results can be considered only as being inconclusive. In addition, the failure to include different orders of materials within each group makes interpretation of the patterns of correlations difficult. But the results tend to support the ATI theory and also suggest that the interactions might be modified as a result of practice over time. Follow-up studies should be conducted to clarify the present results.

E. Summary

The four studies reported in this section were concerned with whether sets of cognitive abilities interact with several methods of teaching vocabulary. The purpose of the first study was to identify cognitive abilities which appear to be related to learning vocabulary by definition and by classification methods. Graduate students studied the learning materials then ordered a set of cognitive abilities according to their belief about the extent to which they were involved. The raters identified six abilities which were common to the materials. However, these abilities were ordered differently for the two sets of materials, thereby suggesting the appropriateness of the learning materials for use in studies of ATI. The second study dealt with determining empirical relationships among the abilities and three methods of teaching vocabulary--synonyms, definitions, and classification. This study was executed on ninth - through twelfth-grade students. The results confirmed the appropriateness of the materials: investigations showed generally that achievement was related differentially to ability depending on the method of instruction, and revealed that raters' perceptions of the important abilities, as determined in the first study, were dissimilar to the obtained empirical relationship. The third study was essentially a replication of the second one, but the subjects were college students, and it generally confirmed the results of the earlier one. The fourth study focussed on changes in the interaction of ability and achievement over an extended period of instruction. The results, although not completely analyzed at the time of preparing this report, suggest that the interactions are modified somewhat systematically over the course of instruction. Because all data were not completely analyzed and because circumstances required that two dissimilar samples of subjects be used, the results of the study merely have suggestive value.

IX. MATHEMATICAL OPERATIONS

The purpose of the studies cited in this section was to determine the replicability of ATI effects through the use of two sets of mathematics materials constructed to be as different as possible in semantic and symbolic content, and ability measures that were as closely related in theory to them as possible. Studies reported in some of the previous sections consisted of one or more pilot studies and a full-scale study. Although the pilot studies were intended primarily to test procedures and materials, the data yielded by these preliminary studies were analyzed and the results were very tentatively regarded as previews of the probable results of the subsequent full-scale study. Because of changes in procedures and materials as a result of the pilot studies and the relatively small number of subjects included in the pilot studies, the similarity of results in the preliminary study and the full-scale study did not agree completely. The studies reported in this section are intended to investigate the stability of the ATI effects over two similar and one somewhat different population.

Each set of instructional materials about mathematical operations was designed to teach the computation of the derivative of an algebraic expression and the multiplication of vectors. For the symbolic material the factor, convergent production of symbolic transformations (NST), was judged to be the most important ability. Conversely the convergent production of semantic transformations (NMT) was judged most relevant for the semantic materials. Tests of eight other factors paired by symbolic-semantic content were judged to be relevant to learning the content and were included in the research.

For every pair of factors, the hypothesis was that the symbolic test would have a significantly larger regression coefficient for the symbolic materials than for the semantic materials. For the semantic test, the regression coefficient for the semantic materials would be larger than that for the symbolic materials.

In the first study a sample of college freshmen was used to determine ATI effects. The second study was an attempt to replicate the results of the first with another sample of freshmen from a different college. The third study was concerned with whether the results of the first two could be extended to high school students.

A. Symbolic and Semantic Learning of Mathematical Operations

Subjects. The students were 71 undergraduates (25 males and 46 females) who were enrolled in an introductory psychology course. They volunteered to participate in the study which required a three-hour block of their time. Each participant was assigned randomly to the semantic or symbolic instructional materials.

Materials. The computation of the derivative of an algebraic expression and the multiplication of vectors comprised the instructional content. These concepts were mediated in semantic and symbolic forms. The materials were compiled in programmed units so that a student could work at his preferred rate. These materials are in Appendix D.

There were three achievement posttests; (a) a test on vector multiplication, (b) a test on the computation of derivatives of certain algebraic expressions, and (c) a combination derivative-vector test. The first two were designed to depend primarily on cognition operation factors. The third was designed to depend primarily on convergent production factors. Copies of these posttests are in Appendix D.

The battery of 10 ability reference tests was assembled in two booklets. The test names and composition of the booklets are given in Table 74. These particular reference tests are described by Guilford and Hoepfner (28) in their factor analytic studies of the structure of intellect.

Procedure. The reference ability tests were administered first. Booklet I was administered, with separate

TABLE 74

The 10 Ability Measures and Their Respective Factors and Reliabilities Used in Mathematical Operations Studies

Tests	Factors
1. Verbal Analogies I	Cognition-Semantic-Relations (CMR)
2. Number Relations	Cognition-Symbolic-Classes (CSC)
3. Word Classification	Cognition-Semantic-Classes (CMC)
4. Gestalt Transformation	Convergent Production-Semantic Transformations (NMT)
5. Sequential Association	Convergent Production-Semantic Implications (NMI)
6. Sign Changes	Convergent Production-Symbolic-Implications (NSI)
7. Seeing Trends II	Cognition-Symbolic-Relations (CSR)
8. Vocabulary Completion	Convergent Production-Semantic Relations (NMR)
9. Correlate Completion II	Convergent Production-Symbolic Relations (NSR)
10. Camouflaged Words	Convergent Production-Symbolic Transformations (NST)

answer sheets, and then Booklet II was administered without separate answer sheets. After a five-minute break, the instructional learning materials were distributed to the students, each student receiving only one version of the material. Each student completed his programmed learning material independently of other students and he noted on his material his starting and finishing time, and the time at which he completed certain parts of the materials.

Analysis. Simple and mutliple regression equations were computed for the semantic and the symbolic groups. Time to complete the learning program and the three posttest achievement scores were the dependent variables. Scores on the ability reference tests were the predictor variables.

Means and standard deviations, for the semantic group and the symbolic group, separately were computed on all aptitude and achievement test scores. Means and standard deviations, for each group separately, were also computed on time measures for completion of the different stages of the learning program.

Results

Each student had scores on ten ability tests, time to complete different stages of the learning program, a derivative test, a vector test and a combination derivative-vector test.

Table 75 contains means and standard deviations on posttests for the groups which studied semantic and symbolic learning materials. The mean difference between groups is not significant at the .05 level.

Table 76 contains a comparison of between-treatment simple regression equations relative to cognition aptitude tests for the semantic and symbolic groups. The dependent variable was the total score on the three posttests. *t*-ratios were computed for individual simple regression coefficients for each group relative to each of the cognition factors.

The tests associated with a cognition-semantic-classes factor and a cognition-semantic-relations factor

TABLE 75

Means and Standard Deviations for Posttest Scores For Semantic and Symbolic Learning Material

Group	Derivative		Vector		Derivative/ Vector		Total	
	M	SD	M	SD	M	SD	M	SD
Semantic	9.86	3.29	10.09	1.60	2.57	1.77	22.54	5.48
Symbolic	11.03	2.93	10.58	1.71	3.28	2.08	24.89	5.46

N = 35 for Semantic Group
N = 36 for Symbolic Group

TABLE 76

Simple Regression Equations of Total Posttests on Cognitive Aptitude Tests

Code	Group	Intercept	Regression Coefficients	Range	Cross-Over	t
CMC	Semantic	13.17	.73	07-18	15.09	2.37
	Symbolic	30.40	-.41	07-18	15.09	- .74
CSC	Semantic	12.83	.48	14-29	<13	2.14
	Symbolic	13.89	.53	13-30	<13	2.84
CMR	Semantic	10.84	.64	10-25	>25	2.68
	Symbolic	18.01	.38	11-24	>25	1.44
CSR	Semantic	13.64	.77	04-17	<03	2.60
	Symbolic	15.05	.81	03-17	<03	2.78

seem to be good predictors of achievement based on the semantic learning material but they are not good predictors of performance based on the symbolic learning material. The two symbolic factors, CSC and CSR, are predictive of performance based on the symbolic material. Only the regression equations associated with the CMC factor have an intersection or crossover point within the range of aptitude scores. The graphs of these two simple regression equations appear in Figure 5.

A comparison of between-treatment simple regression equations relative to convergent production aptitude tests for the semantic and symbolic groups appears in Table 77. The dependent variable was total

TABLE 77

Simple Regression Equations of Total Posttests on Convergent Production Tests

Code	Group	Intercept	Regression Coefficients	Range	Cross-Over	t
NMR	Semantic	- 4.66	.42	40-78	69.21	4.60
	Symbolic	31.34	-.10	45-75	69.21	- .82
NSR	Semantic	12.96	.32	01-39	11.50	3.59
	Symbolic	11.59	.44	00-40	11.50	4.29
NMT	Semantic	13.22	.88	05-16	13.13	3.51
	Symbolic	25.30	-.04	05-16	13.13	- .11
NST	Semantic	18.96	.53	01-15	11.33	1.73
	Symbolic	24.85	.01	01-15	11.33	.02
NMI	Semantic	20.96	.32	01-09	<01	.89
	Symbolic	22.02	.56	01-09	<01	1.50
NSI	Semantic	.47	.64	21-36	>36	2.99
	Symbolic	23.69	.03	23-36	>36	.12

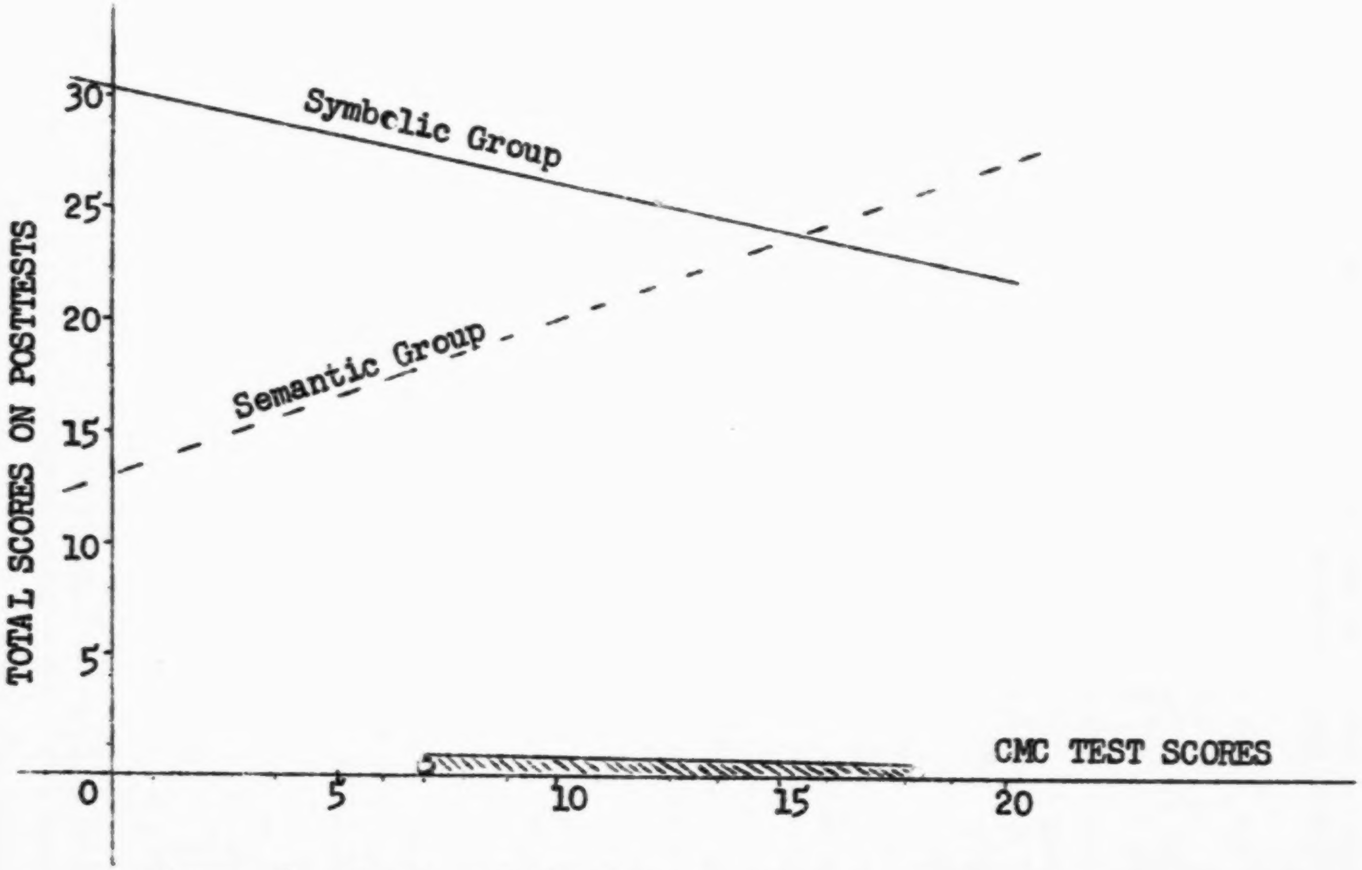

Figure 5. Graphs of simple regression equations for independent variable composed of CMC test scores.

score on the three posttests. t-ratios are also presented for individual regression coefficients for each group relative to each of the convergent production factors.

As Table 77 indicates, the tests associated respectively with a convergent production-semantic-relations (NMR) factor and a convergent production-semantic-transformation (NMT) factor are good predictors of achievement based on the semantic material but they are not relevant as predictors of performance based on symbolic material. The equations associated with the convergent production-symbolic-implication factor do not intersect within the range of aptitude scores. Figures 6, 7 and 8 contain graphs of the regression equations associated with the factors of NMR, NMT, and NSI with crossover points occuring within the range of ability test scores for the first two of these factors. For the NSI factor, the regression line for the symbolic group lies above the regression line for the semantic group. A t-ratio was computed for the difference between regression coefficients (32) for the NSR factor but it was not significant at the .05 level.

Between treatment comparisons of multiple regression equations relative to convergent production-semantic-transformation (NMT) and convergent production-symbolic-transformation (NST) test scores for both the semantic and symbolic groups appear in Tables 78, 79, 80 and 81. The dependent measures, for each of these equations was respectively total scores on posttests, scores on a vector test, scores on a derivative test and scores on a composite derivative-vector test. t-ratios for the individual multiple regression coefficients for each group relative to both factors also appear in the tables. Inspection of the data in Tables 78, 79 and 81 reveals that the test associated with the NMT factor seems to be a good predictor of achievement based on the semantic learning material but it is not a relevant predictor of achievement based on the symbolic learning material. Figure 9 presents graphs of multiple regression equations with the NST and NMT tests as independent variables and with the dependent variable defined by total scores in posttests.

The means and standard deviations of the aptitude test scores for both the semantic and symbolic groups appear in Table 82. There was no significant difference

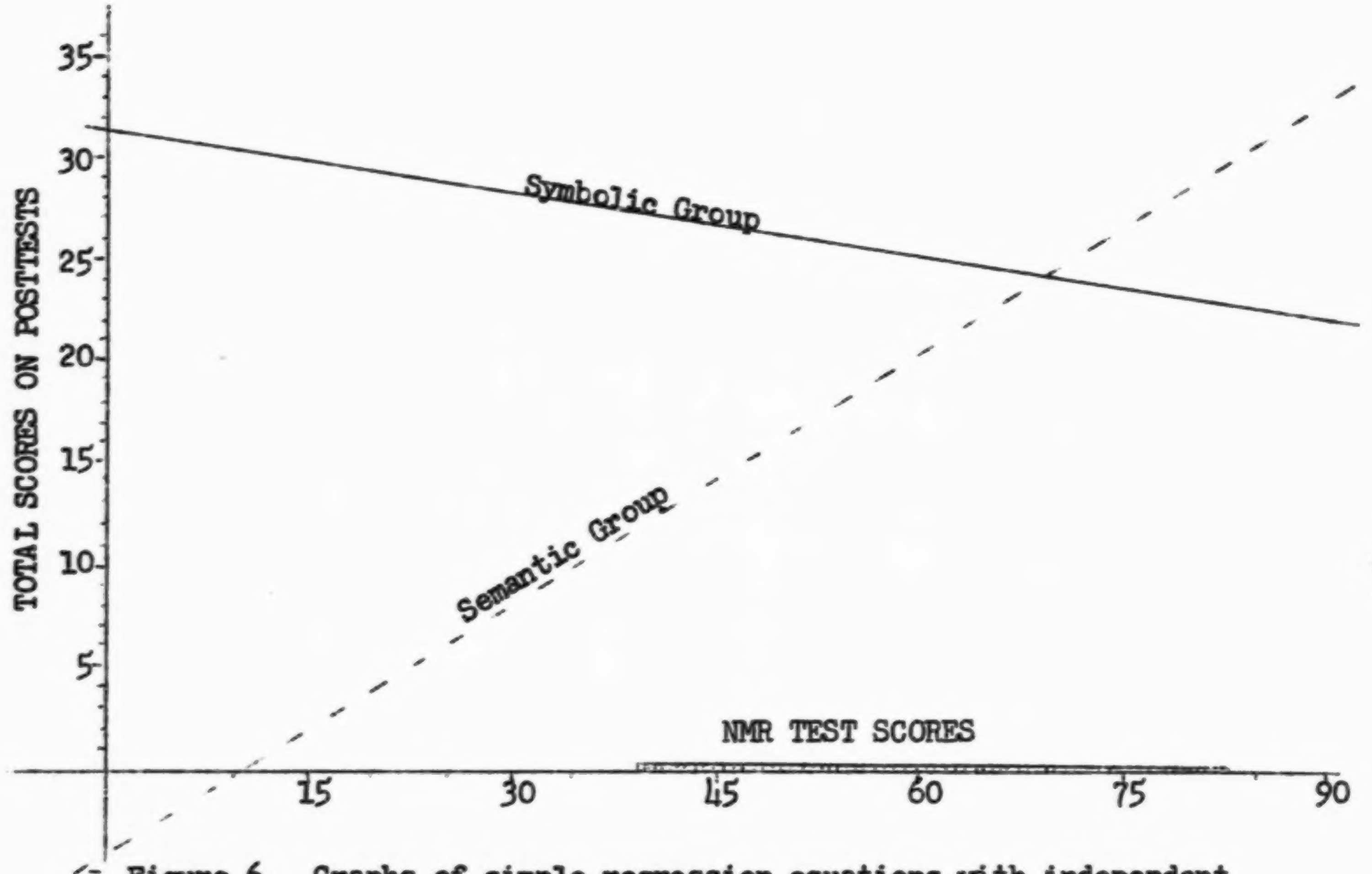

Figure 6. Graphs of simple regression equations with independent variable composed of NMR test scores.

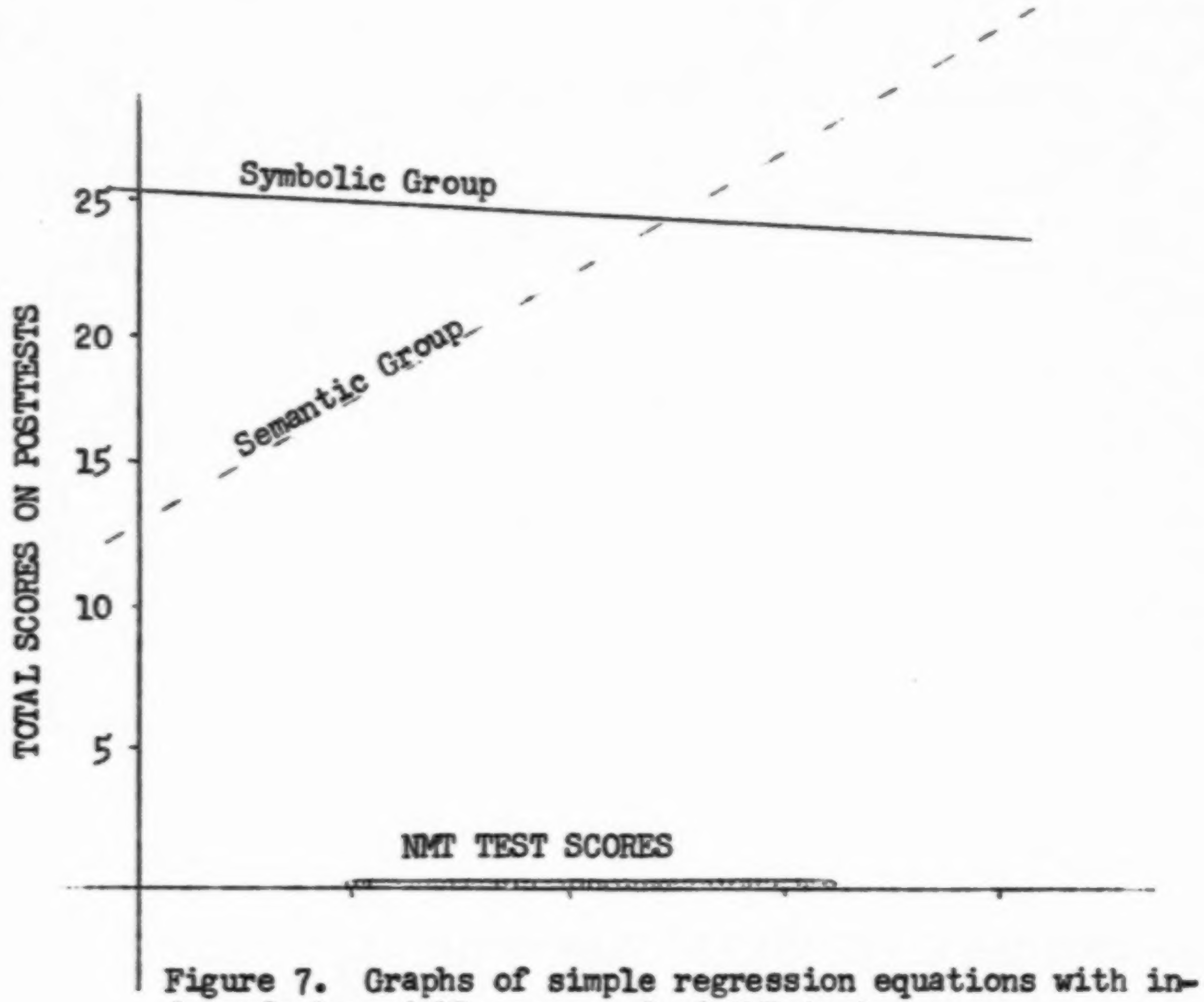

Figure 7. Graphs of simple regression equations with independent variable composed of NMT test scores.

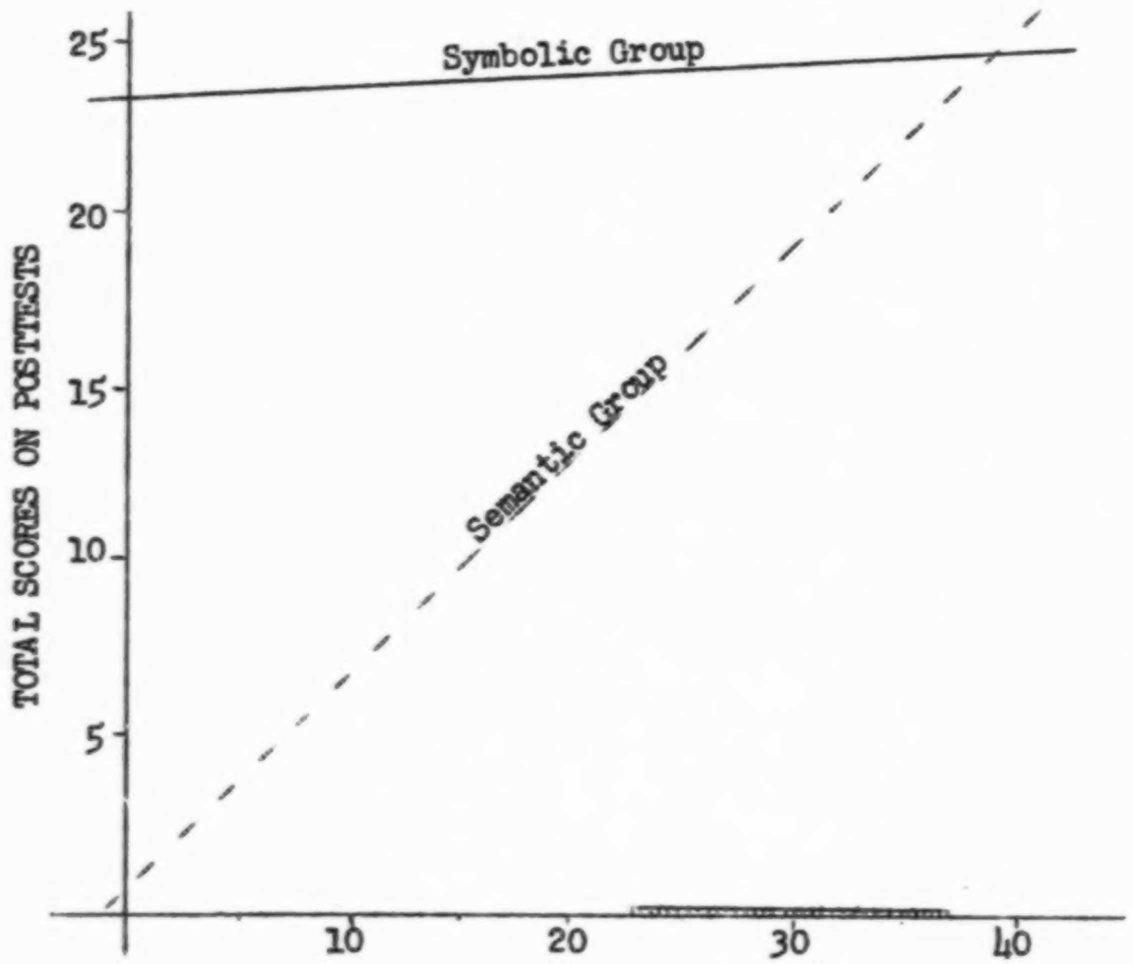

Figure 8. Graphs of simple regression equations with independent variable composed of NSI test scores.

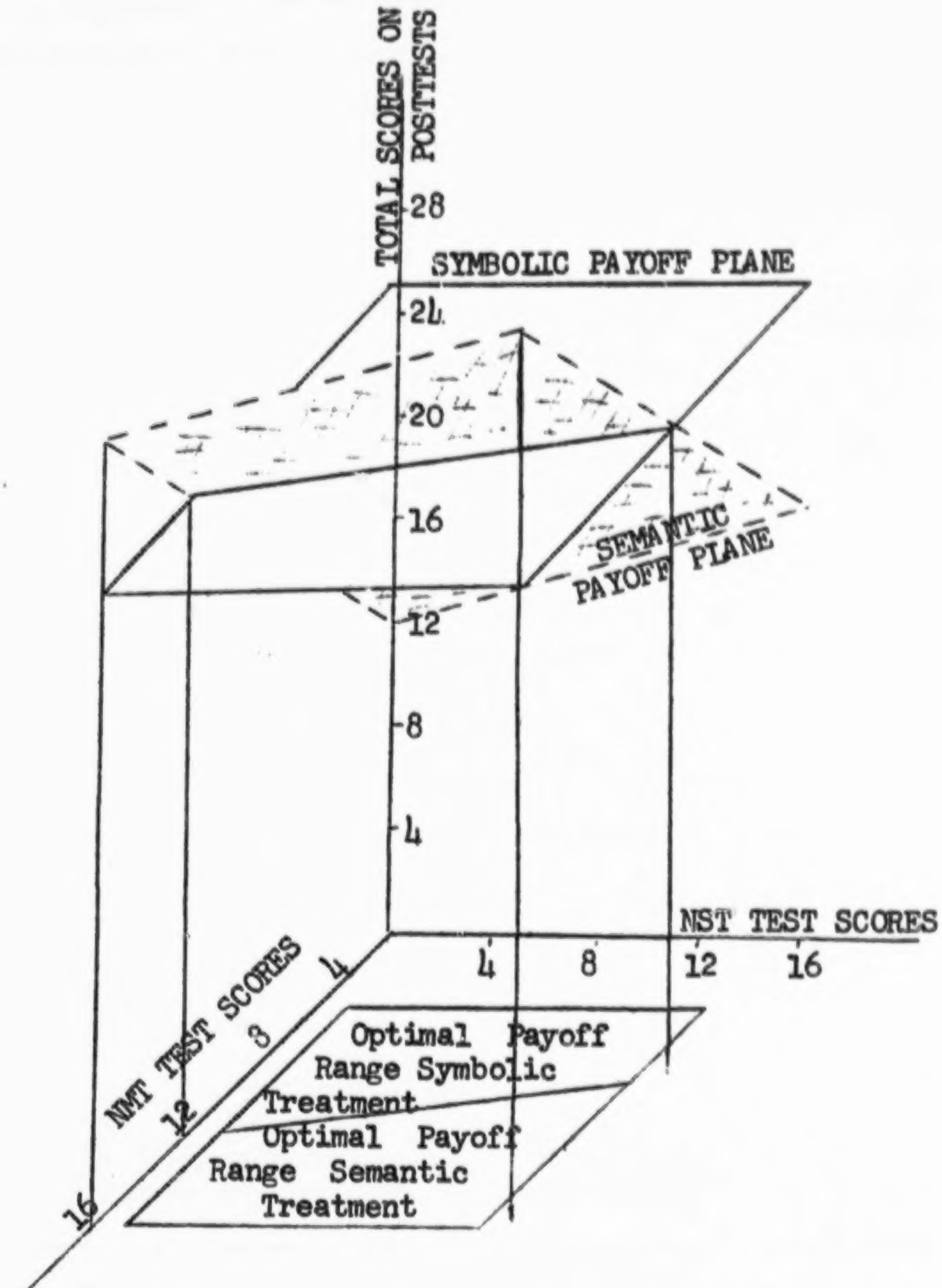

Figure 9. Graphs of multiple regression equations with independent variables composed of NST and NMT test scores

TABLE 78

Multiple Regression Equations of Total Posttests on Transformation Tests

Group	Intercept	(NST) Regression Coefficients	(NMT) Regression Coefficients	(NST) t	(NMT) t
Semantic	12.21	.27	.81	.92	3.06
Symbolic	25.24	.01	-.04	.03	- .12

TABLE 79

Multiple Regression Equations of Vector Tests on Transformation Tests

Group	Intercept	(NST) Regression Coefficients	(NMT) Regression Coefficients	(NST) t	(NMT) t
Semantic	4.60	.20	.37	1.10	2.22
Symbolic	10.40	.06	.02	.38	.09

TABLE 80

Multiple Regression Equations of Derivative Tests on Transformation Tests

Group	Intercept	(NST) Regression Coefficients	(NMT) Regression Coefficients	(NST) t	(NMT) t
Semantic	7.86	.10	.14	1.17	1.67
Symbolic	11.66	-.10	-.03	-1.14	- .26

TABLE 81

Multiple Regression Equations of Derivative/Vector Tests on Transformation Tests

Group	Intercept	(NST) Regression Coefficients	(NMT) Regression Coefficients	(NST) t	(NMT) t
Semantic	- .28	-.04	.30	-.44	3.43
Symbolic	3.18	.05	-.03	.47	- .22

TABLE 82

Means and Standard Deviations of Aptitude Test Scores for the Semantic and Symbolic Groups

Factor	Semantic Group (N=35) M	Semantic Group (N=35) SD	Symbolic Group (N=36) M	Symbolic Group (N=36) SD
CSR	11.57	2.93	12.14	2.90
NMR	64.26	8.07	65.67	7.71
NSR	30.57	9.26	30.17	7.35
NST	6.71	2.97	7.75	3.30
NSI	34.23	3.92	34.33	3.18
NMI	4.97	2.64	5.08	2.41
CMR	18.31	3.63	18.22	3.46
CSC	20.26	4.00	20.58	4.47
CMC	12.89	2.90	13.42	2.12
NMT	10.54	3.23	10.44	2.73

between mean achievement for the two groups at the .05 level for any factor. The intercorrelations of the aptitude tests for the semantic and symbolic groups appear in Tables 83 and 84.

TABLE 83

Intercorrelations of Aptitude Tests

Semantic Group

TEST	CSR	NMR	NSR	NST	NSI	NMI	CMR	CSC	CMC	NMT
CSR		.28	.43	.24	.22	.47	.42	.54	.23	.22
NMR			.44	.60	.26	.04	.44	.30	.45	.40
NSR				.22	.36	.29	.63	.56	.43	.27
NST					.33	.12	.14	.17	.27	.30
NSI						.28	.33	.20	.39	.43
NMI							.32	.47	.18	.15
CMR								.32	.33	.22
CSC									.32	.29
CMC										.35
NMT										

TABLE 84

Intercorrelations of Aptitude Tests

Symbolic Group

TEST	CSR	NMR	NSR	NST	NSI	NMI	CMR	CSC	CMC	NMT
CSR		.23	.60	.25	.17	.01	.34	.43	.32	-.14
NMR			.29	.43	.28	-.06	.27	.15	.36	.11
NSR				.19	.36	.22	.49	.50	-.13	-.01
NST					.22	.01	.30	.17	.33	.13
NSI						.09	.26	.18	-.01	-.28
NMI							-.04	.20	-.12	.10
CMR								.23	-.06	.07
CSC									-.04	.05
CMC										-.19
NMT										

Correlations between ability and achievement tests and total time are contained in Table 85. An inspection of this table seems to indicate that performance on the learning material posttests is not related to time measures.

Intercorrelations of achievement posttest scores and correlations between aptitude tests scores and achievement posttest scores for the semantic and symbolic groups appear in Tables 86 and 87. These indicate the same pattern of intercorrelations for both groups.

Discussion

The results of the study generally support the ATI theory. In seven of ten comparisons the differences between pairs of simple regression coefficients were in the predicted direction. Reversals of the predicted direction were found for factors NST, NSI, and NMI. The multiple regression equations also support the theory but less satisfactorily. NMT was a significant predictor of all but one criterion for the semantic group and was never significant for the symbolic group. NST had no significant coefficients in any of the multiple regression equations.

B. Symbolic and Semantic Learning of Mathematical Operations: Replication with a Similar Population

The purpose of this study was to test further the results of the previous study by essentially replicating the earlier study.

Subjects. The students were 73 undergraduates (38 males and 35 females) who were enrolled in a general psychology course. They volunteered a three-hour block of time for the study. Students were assigned randomly to the semantic and symbolic instructional materials.

Materials. The materials were the same as those used in the earlier study.

TABLE 85

Correlations Between Ability and Achievement Tests and Total Time

Test	Time on Learning Program Semantic Group	 Symbolic Group
Derivative	.01	.17
Vector	.02	.13
Derivative/ Vector	-.04	-.10
Total	.00	.09
CSR	-.27	-.14
NMR	.13	-.39
NSR	-.15	.10
NST	.19	-.21
NSI	.08	-.29
NMI	-.32	.24
CMR	-.07	-.03
CSC	-.47	-.25
CMC	-.05	-.40
NMT	-.07	.13

TABLE 86

Correlations Between Aptitude Tests and Achievement Posttests

Semantic Group

Test	Derivative	Vector	Derivative/ Vector	Total
CSR	.36	.19	.39	.41
NMR	.57	.35	.56	.62
NSR	.42	.48	.43	.53
NST	.29	.28	.09	.29
NSI	.46	.34	.26	.46
NMI	.21	.10	-.01	.15
CMR	.29	.36	.44	.42
CSC	.28	.34	.27	.35
CMC	.31	.26	.37	.38
NMT	.42	.34	.52	.52
Derivative		.35	.59	.89
Vector			.53	.67
Derivative/ Vector				.83
Total				

TABLE 87

Correlations Between Aptitude Tests and Achievement Posttests

Symbolic Group

Test	Derivative	Vector	Derivative/ Vector	Total
CSR	.37	.29	.37	.43
NMR	-.08	-.22	-.07	-.12
NSR	.62	.38	.37	.59
NST	.07	-.20	.07	.00
NSI	.07	-.06	.00	.02
NMI	.26	.21	.11	.25
CMR	.27	-.05	.29	.24
CSC	.47	.26	.26	.44
CMC	-.20	-.10	-.05	-.16
NMT	.02	-.07	-.03	-.02
Derivative		.39	.65	.91
Vector			.30	.64
Derivative/ Vector				.82
Total				

Procedure. The procedure was the same used in the earlier study.

Analysis. The analysis was the same as that used in the previous study.

Results

Means and standard deviations for posttests on the semantic and symbolic learning material appear in Table 88. It shows a significant difference between mean achievement for the two groups at the .05 level, the symbolic group being consistently higher.

TABLE 88

Means and Standard Deviations for Posttest Scores for Semantic and Symbolic Learning Material

Group	Derivative		Vector		Derivative/ Vector		Total	
	M	SD	M	SD	M	SD	M	SD
Semantic	8.14	4.05	8.84	3.23	1.46	1.37	18.43	6.64
Symbolic	10.25	3.97	10.53	1.32	2.78	2.11	23.53	6.17

N = 37 for Semantic Group
N = 36 for Symbolic Group

A comparison of between-treatment simple regression equations relative to cognition aptitude tests for the semantic and symbolic groups appear in Table 89. The dependent variable was total achievement scores on the three posttests. t-ratios are also shown for individual simple regression coefficients for each group relative to each of the cognition factors. The data reveals that the tests representing the CMC, CMR and CSR factors seem

TABLE 89

Simple Regression Equations of Total Posttests on Cognitive Aptitude Tests

Code	Group	Intercept	Regression Coefficients	Range	Cross-Over	t
CMC	Semantic	8.36	.86	07-16	15.61	2.42
	Symbolic	29.43	-.49	07-18	15.61	-1.27
CSC	Semantic	12.91	.32	10-26	<10	1.27
	Symbolic	15.43	.42	11-28	<10	2.05
CMR	Semantic	7.55	.68	08-25	>25	2.75
	Symbolic	16.35	.41	11-24	>25	1.58
CSR	Semantic	5.69	1.12	07-15	>16	3.01
	Symbolic	19.20	.39	03-16	>16	1.15

to be good predictors of achievement based on the semantic learning material but not of achievement based on the symbolic material. Only the regression equations associated with the CMC factor have an intersection or crossover point within the range of aptitude scores. For both the CMR and CSR factors the symbolic regression equation lies above the semantic regression equation throughout the range of aptitude scores. The graphs of the regression equations associated with the CMC factor having a crossover point occuring within the range of aptitude scores appear in Figure 10.

A comparison of between-treatment simple regression equations relative to convergent production aptitude tests for the semantic and symbolic groups appears in Table 90. The dependent variable was total achievement scores on the three posttests. *t*-ratios are shown for individual regression coefficients for each group relative to each of the convergent production factors.

The tests associated with the NMR, NMI and the NSI factors seem to be good predictors of achievement

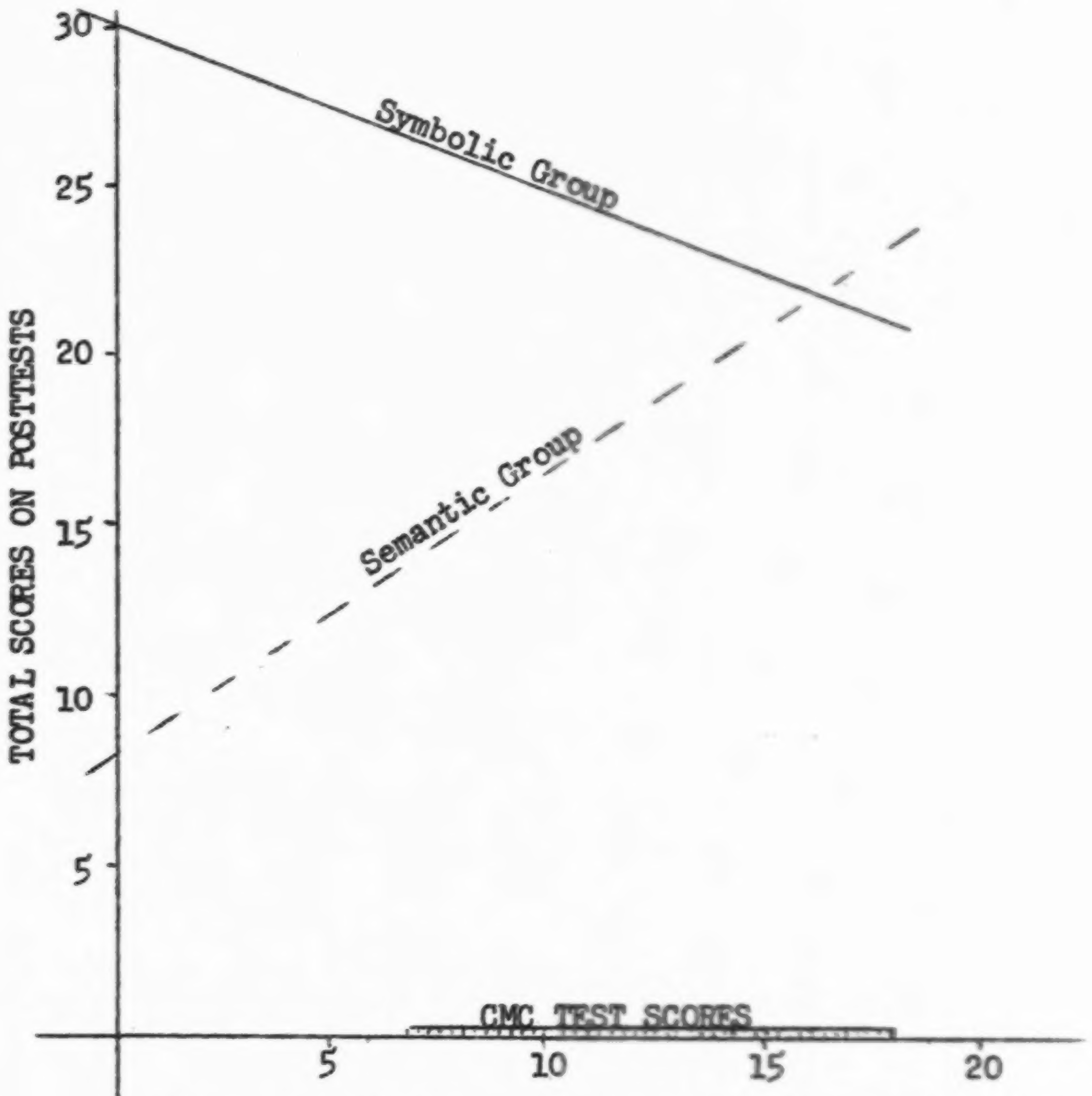

Figure 10. Graphs of simple regression equations for independent variable composed of CMC test scores.

TABLE 90

Simple Regression Equations of Total Posttests on Convergent Production Tests

Code	Group	Intercept	Regression Coefficients	Range	Cross-Over	t
NMR	Semantic	- 7.99	.42	41-78	74.56	4.57
	Symbolic	24.82	-.02	39-74	74.56	- .19
NSR	Semantic	6.40	.44	16-39	>39	2.96
	Symbolic	15.03	.29	17-39	>39	2.13
NMT	Semantic	15.69	.26	05-16	>16	.64
	Symbolic	22.62	.08	02-16	>16	.25
NST	Semantic	13.05	.81	02-15	11.29	1.94
	Symbolic	25.70	-.31	03-15	11.29	.90
NMI	Semantic	13.76	1.46	01-07	8.47	3.02
	Symbolic	20.96	.61	01-09	8.47	1.59
NSI	Semantic	- 4.63	.67	24-36	25.02	2.06
	Symbolic	-15.14	1.09	32-36	25.02	1.39

based on the semantic material but are not relevant predictors of achievement based on the symbolic material. Graphs of the regression equations associated with these factors appear in Figures 11, 12, and 13. The crossover points occur within the range of ability scores.

Between-treatment comparison of multiple regression equations relative to NMT and NST test scores for both the semantic and symbolic groups appear in Tables 91, 92, 93 and 94. The dependent variables for each of these equations were total scores on posttests, scores on a vector test, scores on a derivative test and scores on a composite derivative-vector test, respectively. *t*-ratios were calculated for the individual multiple regression coefficients for each group relative to both the NMT factor and the NST factor.

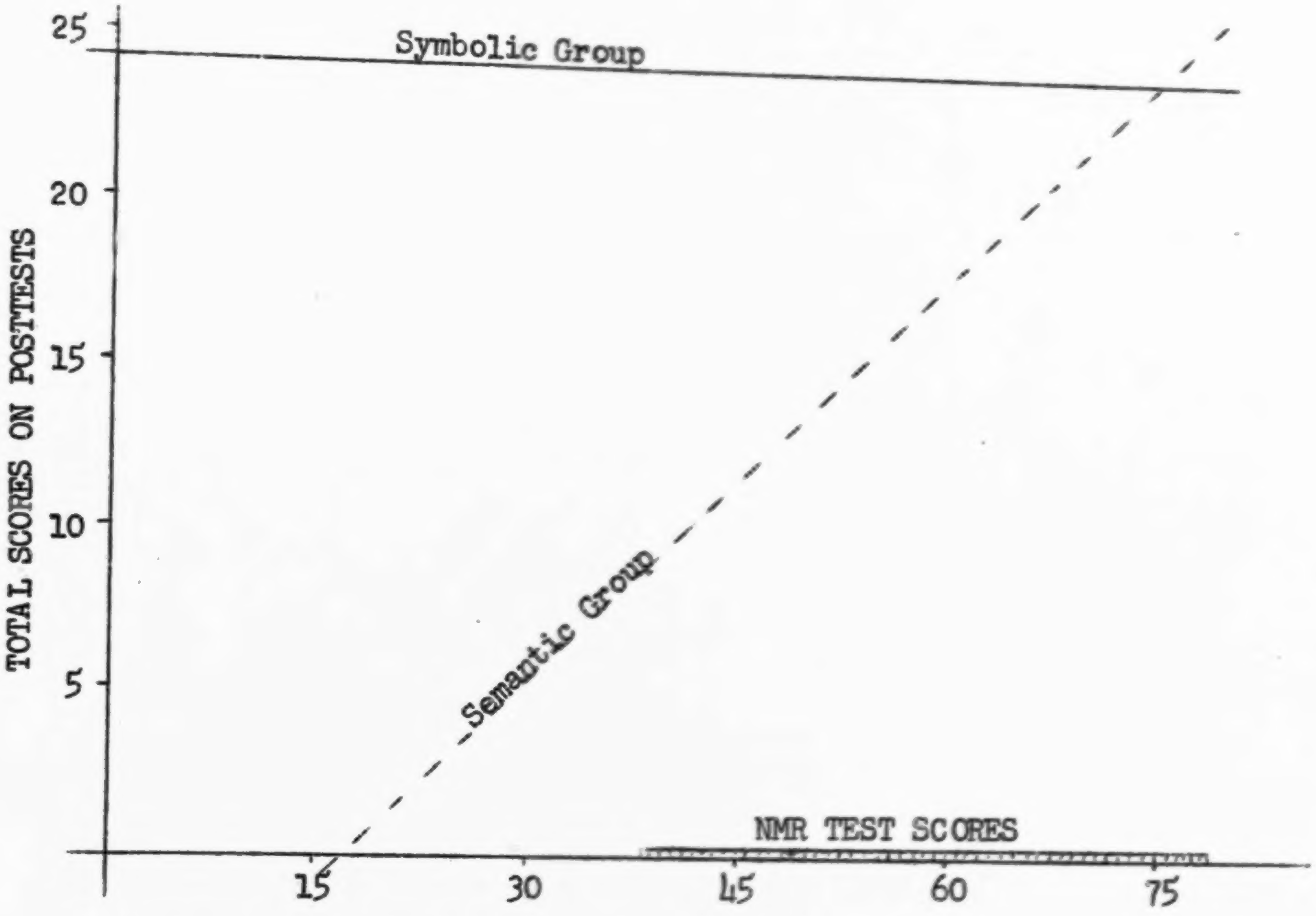

Figure 11. Graphs of simple regression equations for independent variable composed of NMR test scores.

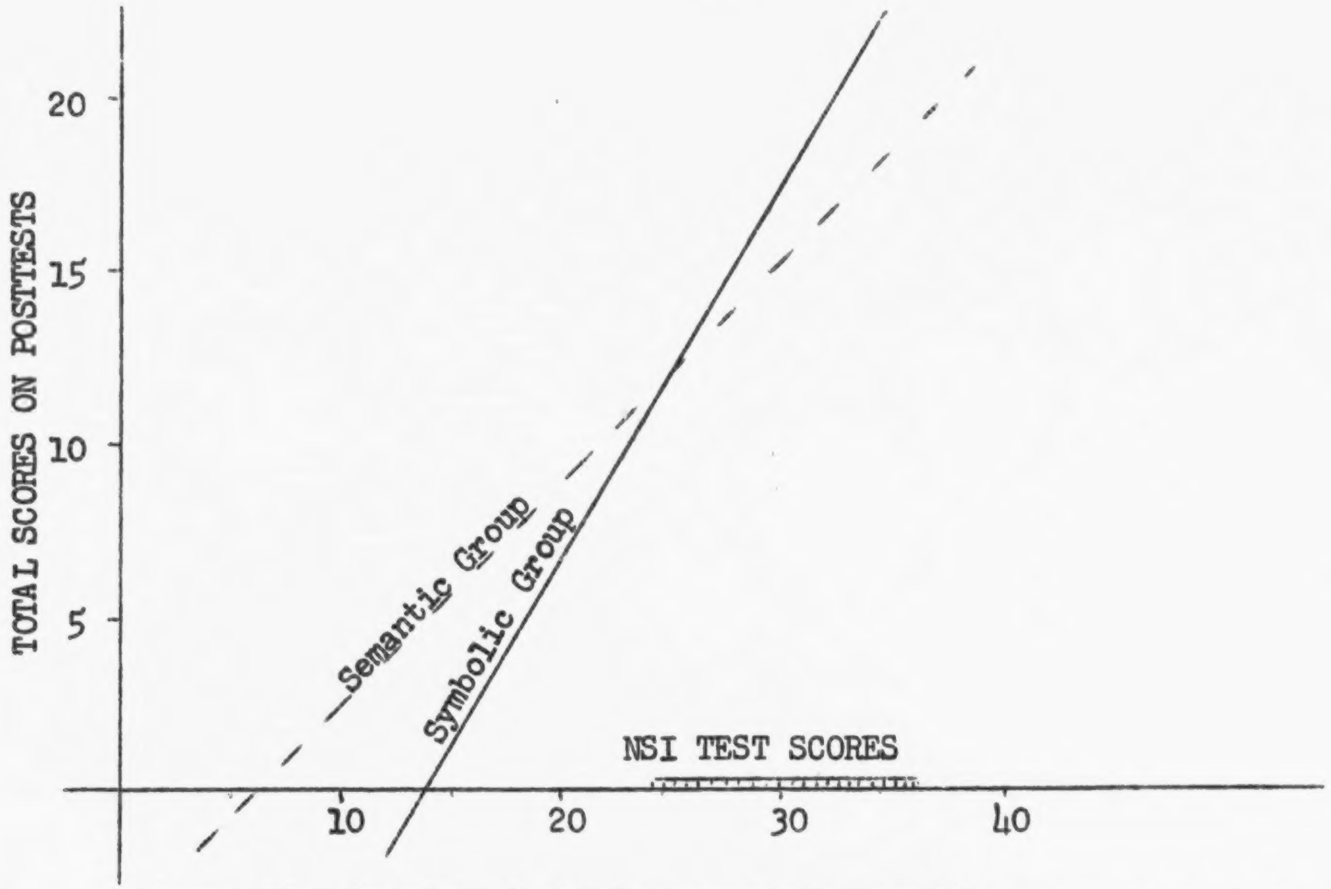

Figure 12. Graphs of simple regression equations for independent variable composed of NSI test scores.

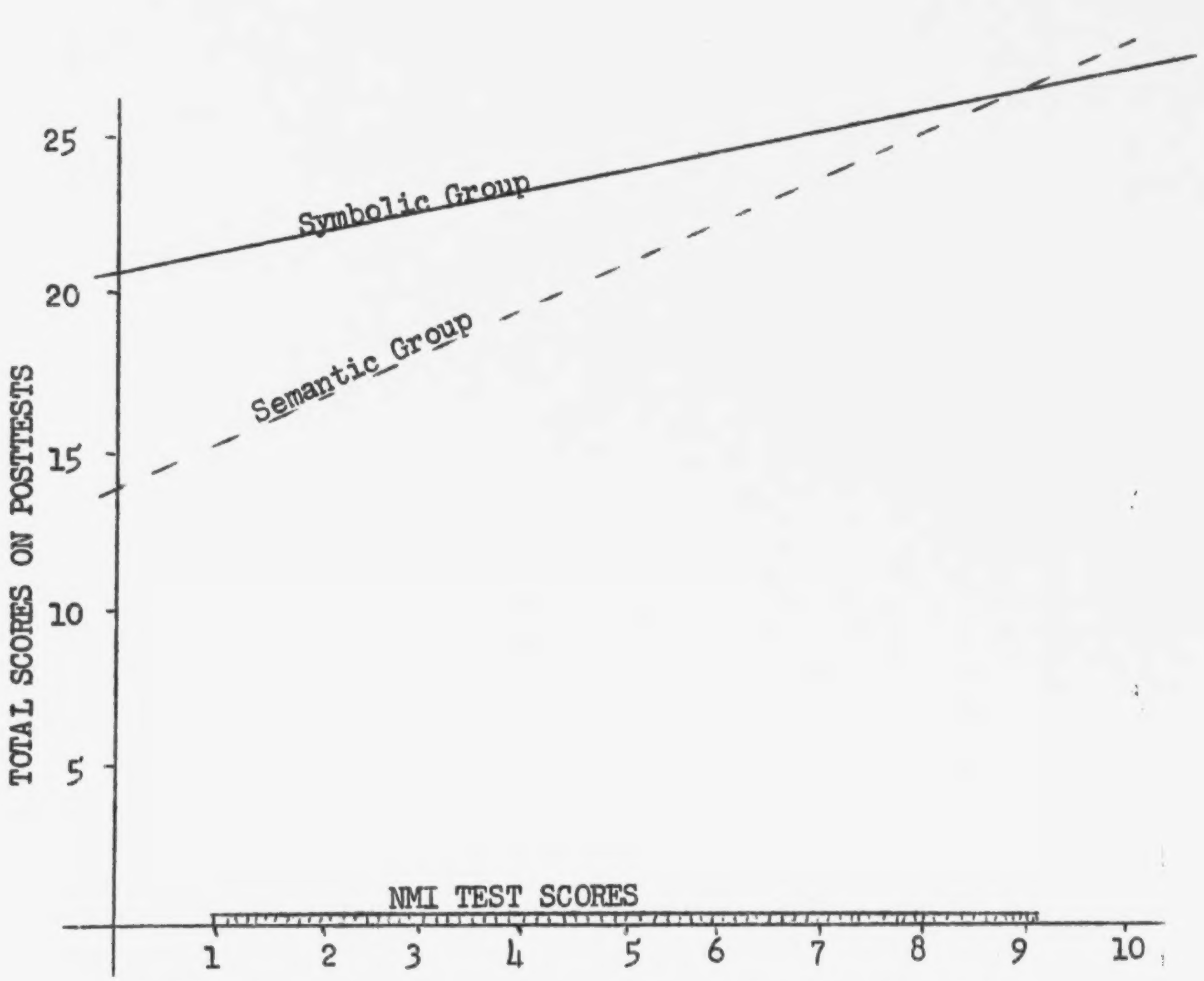

Figure 13. Graphs of simple regression equations for independent variable composed of NMI test scores.

TABLE 91

Multiple Regression Equations of Total Posttests on Transformation Tests

Group	Intercept	(NST) Regression Coefficients	(NMT) Regression Coefficients	(NST) t	(NMT) t
Semantic	10.67	.80	.24	1.89	.59
Symbolic	24.31	-.34	.14	- .96	.43

TABLE 92

Multiple Regression Equations of Vector Tests on Transformation Tests

Group	Intercept	(NST) Regression Coefficients	(NMT) Regression Coefficients	(NST) t	(NMT) t
Semantic	5.02	.32	.16	1.53	.84
Symbolic	10.14	-.09	.09	-1.24	1.34

TABLE 93

Multiple Regression Equations of Derivative Tests on Transformation Tests

Group	Intercept	(NST) Regression Coefficients	(NMT) Regression Coefficients	(NST) t	(NMT) t
Semantic	5.66	.38	-.01	1.44	- .02
Symbolic	10.32	-.12	.07	- .52	.32

TABLE 94

Multiple Regression Equations of
Derivative/Vector Tests on
Transformation Tests

Group	Intercept	(NST) Regression Coefficients	(NMT) Regression Coefficients	(NST) t	(NMT) t
Semantic	5.03	.32	.16	1.53	.84
Symbolic	3.78	-.13	-.01	-1.08	-.07

Means and standard deviations of the aptitude test scores for both the semantic and symbolic groups appear in Table 95. There was no significant difference between achievement for the two groups at the .05 level on any factor. Intercorrelations of the aptitude tests for the semantic and symbolic groups are contained in Tables 96 and 97.

Correlations between ability and achievement tests and total time appear in Table 98. These indicate that performance on posttest is not related to learning time for the semantic material but is related to time measures for the symbolic material. Intercorrelations of achievement posttest scores and correlations between aptitude test scores and achievement posttest scores for the semantic and symbolic groups appear in Tables 99 and 100.

Discussion

The results for the cognitive factors of this study are in good agreement with those of the first study. For three of the four cognition tests, the direction of the differences between the simple regression coefficients was the same as in the first study and was in the predicted direction. The direction of the difference between the coefficients for the test of factor CSR was not in the predicted direction. In addition the crossover points for the pairs of regression

TABLE 95

Means and Standard Deviations of Aptitude Test Scores for the Semantic and Symbolic Groups

	Semantic Group (N=37)		Symbolic Group (N=36)	
Factor	M	SD	M	SD
CSR	11.38	2.69	11.22	3.09
NMR	61.78	9.50	61.08	9.28
NSR	27.16	6.70	29.03	7.23
NST	6.62	2.54	7.03	3.05
NSI	34.24	3.24	35.33	1.31
NMI	3.19	2.07	4.19	2.65
CMR	16.11	4.14	17.44	3.93
CSC	17.30	4.36	19.14	4.84
CMC	11.68	2.92	12.08	2.68
NMT	10.38	2.71	11.06	3.22

TABLE 96

Intercorrelations of Aptitude Tests

Semantic Group

TEST	CSR	NMR	NSR	NST	NSI	NMI	CMR	CSC	CMC	NMT
CSR		.52	.49	.36	.19	.40	.48	.32	.39	-.22
NMR			.52	.48	.38	.36	.40	.50	.54	.22
NSR				.18	.15	.43	.29	.46	.53	.04
NST					.04	.14	.28	.29	.20	.04
NSI						.14	.19	.15	.22	.21
NMI							.54	.35	.38	.16
CMR								.30	.21	.20
CSC									.42	.19
CMC										.29
NMT										

TABLE 97

Intercorrelations of Aptitude Tests

Symbolic Group

TEST	CSR	NMR	NSR	NST	NSI	NMI	CMR	CSC	CMC	NMT
CSR		.31	.49	.25	.31	.32	.36	.13	.22	.28
NMR			.26	.39	.13	.04	.49	.31	.42	.32
NSR				.13	.34	.62	.21	.30	.13	.20
NST					-.06	.24	.30	.22	.20	.19
NSI						.07	.06	.33	.38	.00
NMI							.25	.04	.06	.16
CMR								.44	.26	.46
CSC									.03	.13
CMC										.16
NMT										

TABLE 98

Correlations Between Ability and Achievement Tests and Total Time

Test	Time on Learning Program Semantic Group	 Symbolic Group
Derivative	-.02	.41
Vector	.19	.40
Derivative/ Vector	.22	.31
Total	.12	.45
CSR	.04	.16
NMR	-.09	.03
NSR	-.19	.29
NST	-.15	.14
NSI	.03	-.12
NMI	-.14	.23
CMR	-.04	.13
CSC	-.13	-.13
CMC	-.36	-.21
NMT	-.17	.02

TABLE 99

Correlations Between Aptitude Tests and Achievement Posttests

Semantic Group

Test	Derivative	Vector	Derivative/ Vector	Total
CSR	.39	.27	.43	.45
NMR	.57	.33	.49	.61
NSR	.37	.28	.43	.45
NST	.24	.26	.20	.31
NSI	.40	.11	.14	.33
NMI	.30	.41	.37	.46
CMR	.25	.46	.22	.42
CSC	.15	.08	.37	.21
CMC	.40	.12	.37	.38
NMT	.01	.15	.16	.11
Derivative		.22	.56	.83
Vector			.40	.70
Derivative/ Vector				.74
Total				

TABLE 100

Correlations Between Aptitude Tests and Achievement Posttests

Symbolic Group

Test	Derivative	Vector	Derivative/ Vector	Total
CSR	.14	.28	.13	.19
NMR	.06	.04	-.22	-.03
NSR	.38	.13	.23	.34
NST	-.08	-.17	-.19	-.15
NSI	.31	-.17	.18	.23
NMI	.18	.03	.40	.26
CMR	.27	.03	.23	.26
CSC	.45	-.14	.20	.33
CMC	-.14	-.13	-.27	-.21
NMT	.04	.19	-.05	.04
Derivative		.35	.67	.95
Vector			.23	.51
Derivative/ Vector				.83
Total				

coefficients of all cognition factors except CSR were very similar to those of the first study. The reasons for the reversal of CSR are unknown.

The results for the convergent production factors were not in good agreement with those of the first study. Three of the six factors--NSR, NMI and NSI--showed directions of differences between the magnitudes of the simple regression coefficients that were the opposite of those of the first study. In the case of NSI, however, only the regression coefficient for the semantic group was significantly different from zero in both studies. While neither coefficient for NMI was significantly different from zero in the first study, its coefficient in the second study for the semantic group was significant. This finding is consistent with the predicted difference. The crossover points for two of the factors for which agreement of regression coefficients was found are also in good agreement. These factors are NMR and NST.

None of the multiple regression equations was similar to those of the first study. In every instance the regression coefficient for NST was higher for the semantic group than was the coefficient for NMT. In the symbolic group the results were reversed. None of the _t_-ratios for the regression coefficients of either factor for either group was significant in this study.

Taken together the results from the two studies indicate that the differences between the simple regression coefficients are fairly stable. In addition, these differences generally behave in the manner suggested by the ATI theory. The differences between multiple regression coefficients are unstable and do not conform to ATI theory. The use of larger groups might reduce the instability of the multiple regressions.

C. Symbolic and Semantic Learning of Mathematical Operations: Replication with a Dissimilar Population

The main purpose of this study was to test further the findings of aptitude achievement interaction as reported in Studies I and II by the use of a sample drawn from a different type of population.

Subjects. The subjects were 177 tenth-grade students (79 males and 98 females). The study extended over three consecutive days for one hour each morning. The ability reference tests were administered on the first two days. The learning programs and posttests were administered on the last day. Students were assigned randomly to the semantic and symbolic instructional materials. 91 were in the semantic group and 86 in the symbolic group.

Materials. The materials were the same as those used in the previous studies.

Procedure. The procedure used was the same as that used in the previous studies except for spreading the study over three days.

Analysis. The method analysis was the same as that used in the previous studies.

Results

Means and standard deviations for posttest performance based on the semantic and symbolic learning materials appear in Table 101. A significant difference

TABLE 101

Means and Standard Deviations for Posttest Scores For Semantic and Symbolic Learning Material

Group	Derivative		Vector		Derivative/ Vector		Total	
	M	SD	M	SD	M	SD	M	SD
Semantic	5.47	4.04	6.25	4.51	.71	1.14	12.42	8.26
Symbolic	4.10	4.02	8.24	3.44	.72	1.27	13.03	7.05

N = 91 for Semantic Group
N = 86 for Symbolic Group

occurs between mean achievement for the two groups on the derivative test and the vector test at the .05 level.

A comparison of between-treatment simple regression equations relative to cognition aptitude tests for the semantic and symbolic groups appears in Table 102. The dependent variable is total score on the three posttests. *t*-ratios are also shown for individual simple regression coefficients for each group relative to each of the cognition factors.

TABLE 102

Simple Regression Equations of Total Posttests on Cognitive Aptitude Tests

Code	Group	Intercept	Regression Coefficients	Range	Cross-Over	t
CMC	Semantic	4.45	.76			2.35
	Symbolic	2.14	1.08	05-16	7.21	3.88
CSC	Semantic	- .89	.92			6.64
	Symbolic	4.49	.60	02-28	11.25	4.54
CMR	Semantic	- .09	.97			4.73
	Symbolic	7.25	.49	06-24	15.29	2.76
CSR	Semantic	4.11	1.08			5.34
	Symbolic	6.52	.97	00-15	>15	4.80

Three tests associated with the cognition aptitudes factors (CMC, CSC, CMR) seem to be relevant predictors for both the semantic and the symbolic learning material. Graphs of these simple regression equations appear in Figures 14 through 16.

A comparison of between-treatment simple regression equations relative to convergent production aptitude tests for the semantic and symbolic groups appears in Table 103. The dependent variable was total

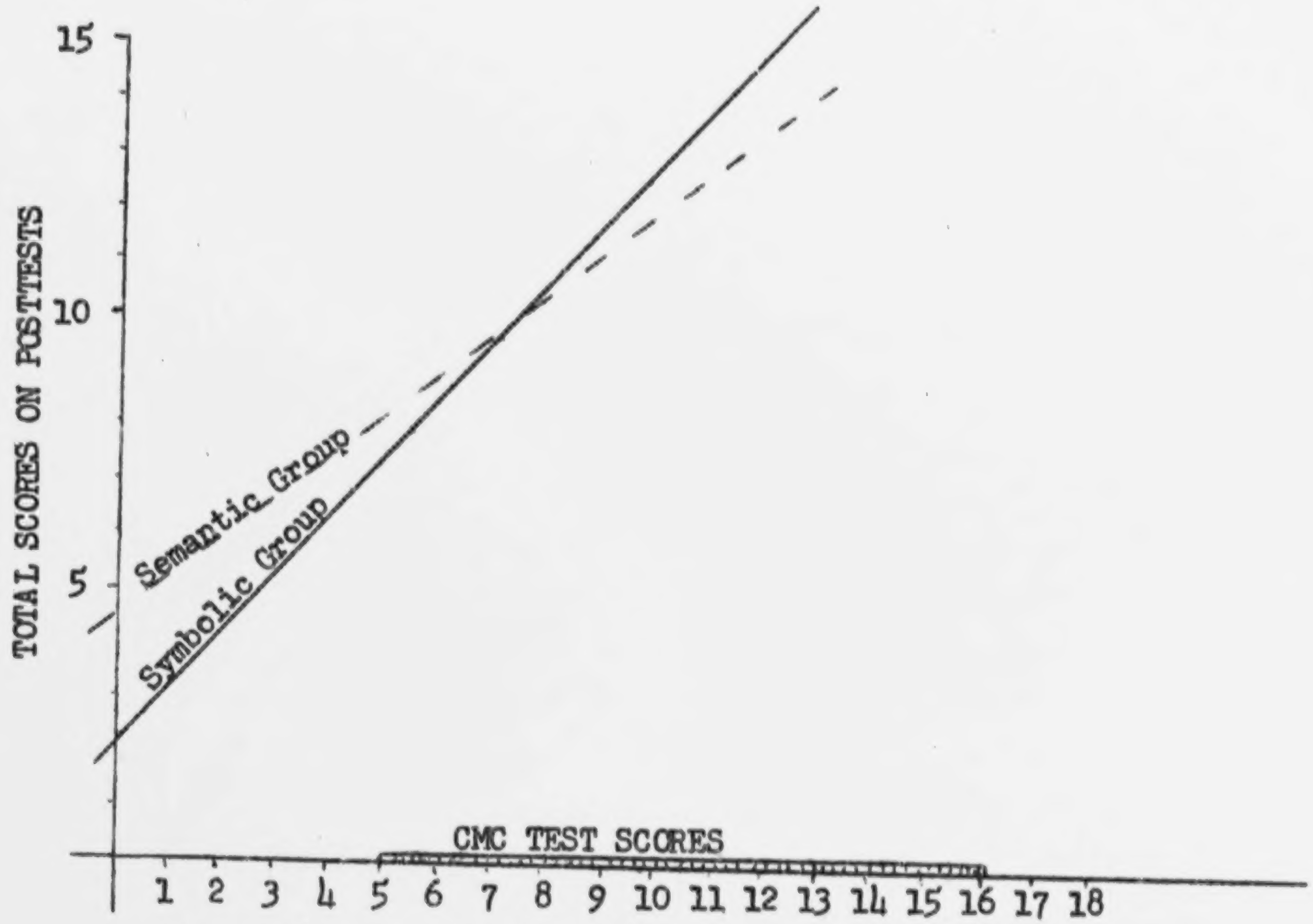

Figure 14. Graphs of simple regression equations for independent variable composed of CMC test scores.

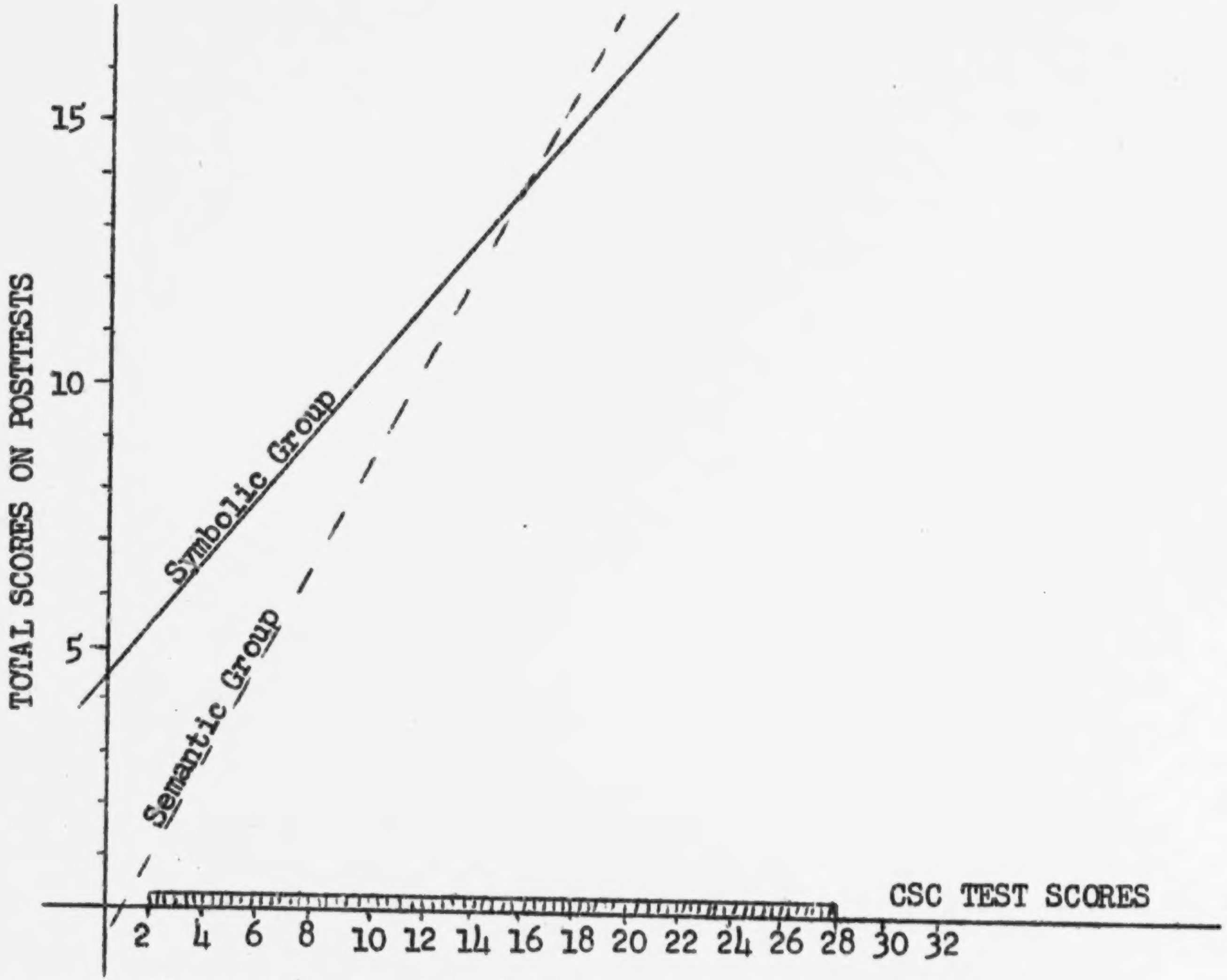

Figure 15. Graphs of simple regression equations for independent variable composed of CSC test scores.

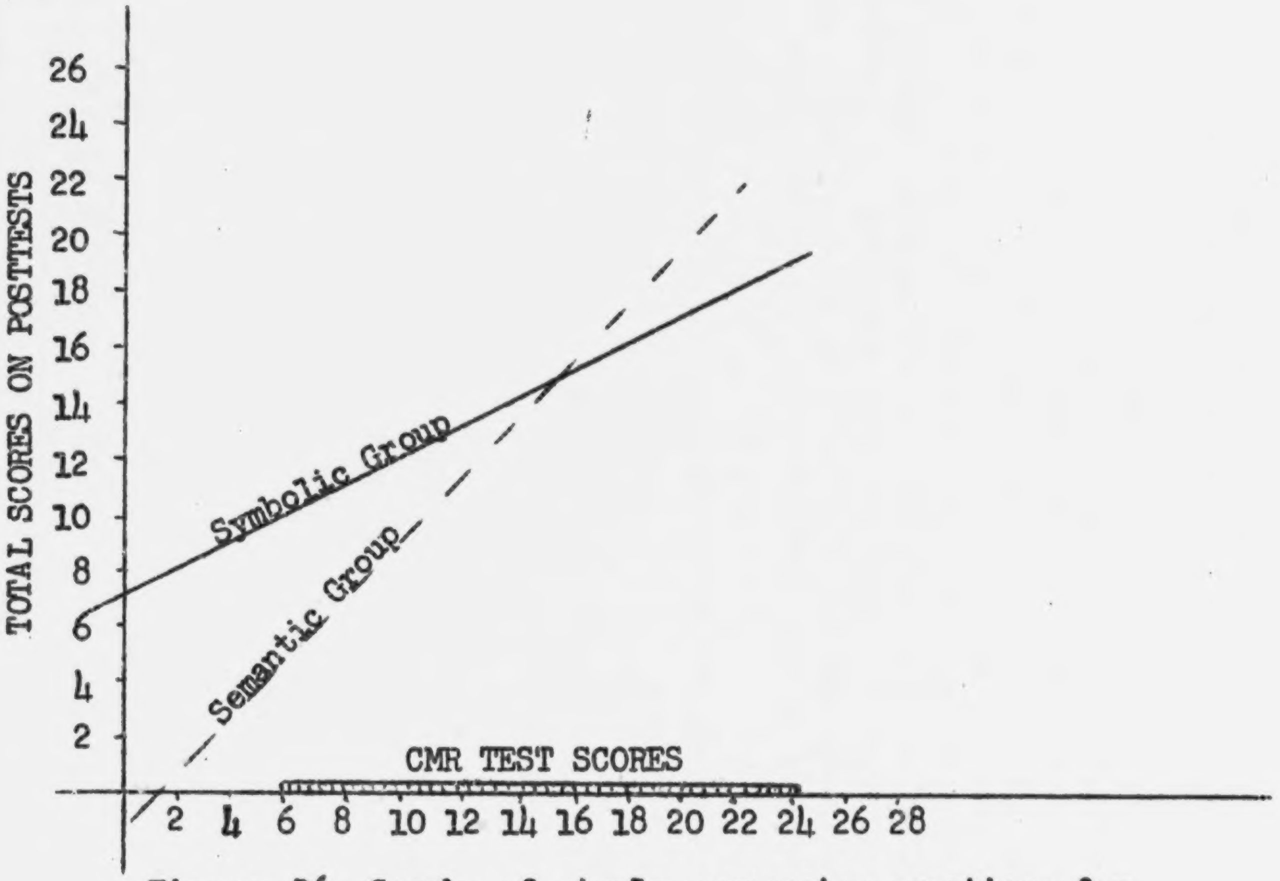

Figure 16. Graphs of simple regression equations for independent variable composed of CMR test scores.

TABLE 103

Simple Regression Equations of Total Posttests on Convergent Production Tests

Code	Group	Intercept	Regression Coefficients	Range	Cross-Over	t
NMR	Semantic	- 1.88	.30			4.54
	Symbolic	5.81	.15	08-73	51.27	2.49
NSR	Semantic	2.30	.46			7.81
	Symbolic	8.79	.23	00-38	28.22	3.83
NMT	Semantic	6.88	.77			2.75
	Symbolic	7.11	.96	00-15	00	4.05
NST	Semantic	9.01	.64			2.06
	Symbolic	10.86	.42	00-15	8.41	1.64
NMI	Semantic	7.61	1.63			4.25
	Symbolic	10.13	1.03	00-08	4.20	2.48
NSI	Semantic	- 3.90	.50			3.41
	Symbolic	8.78	.13	01-36	34.54	1.10

achievement on the three posttests. t-ratios were calculated for individual regression coefficients for each group relative to each of the convergent production factors. Graphs of the regression equations associated with the factors of NMR and NSR appear in Figures 17 and 18. t-ratios were computed for the differences between regression coefficients (32) for the two groups with respect to each of the aptitude factors. Only the difference between regression coefficients for the NSR factor was significantly different at the .05 level.

Between-treatment comparison of multiple regression equations relative to NMT and NST test scores for both the semantic and symbolic groups appear in Tables 104 through 107. The dependent variables were total scores on posttests, scores on a vector test, scores on a derivative test and scores on a composite derivative-vector test, respectively.

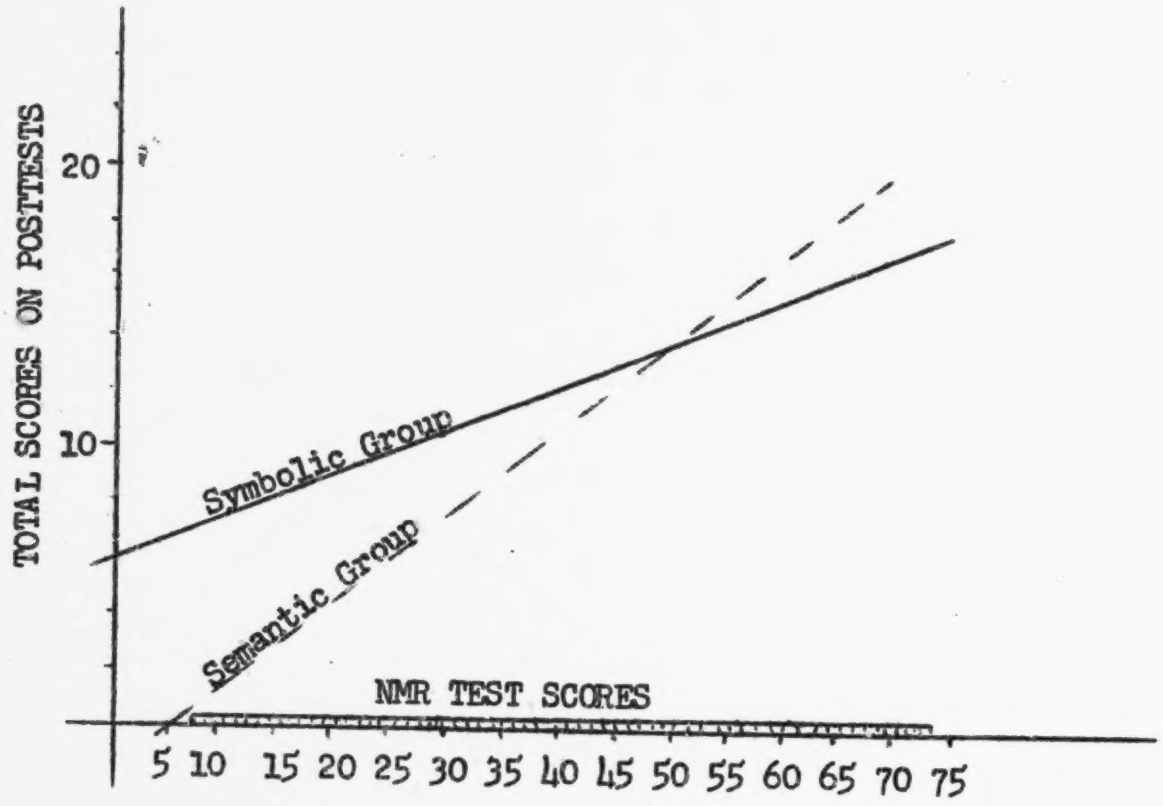

Figure 17. Graphs of simple regression equations for independent variable composed of NMR test scores.

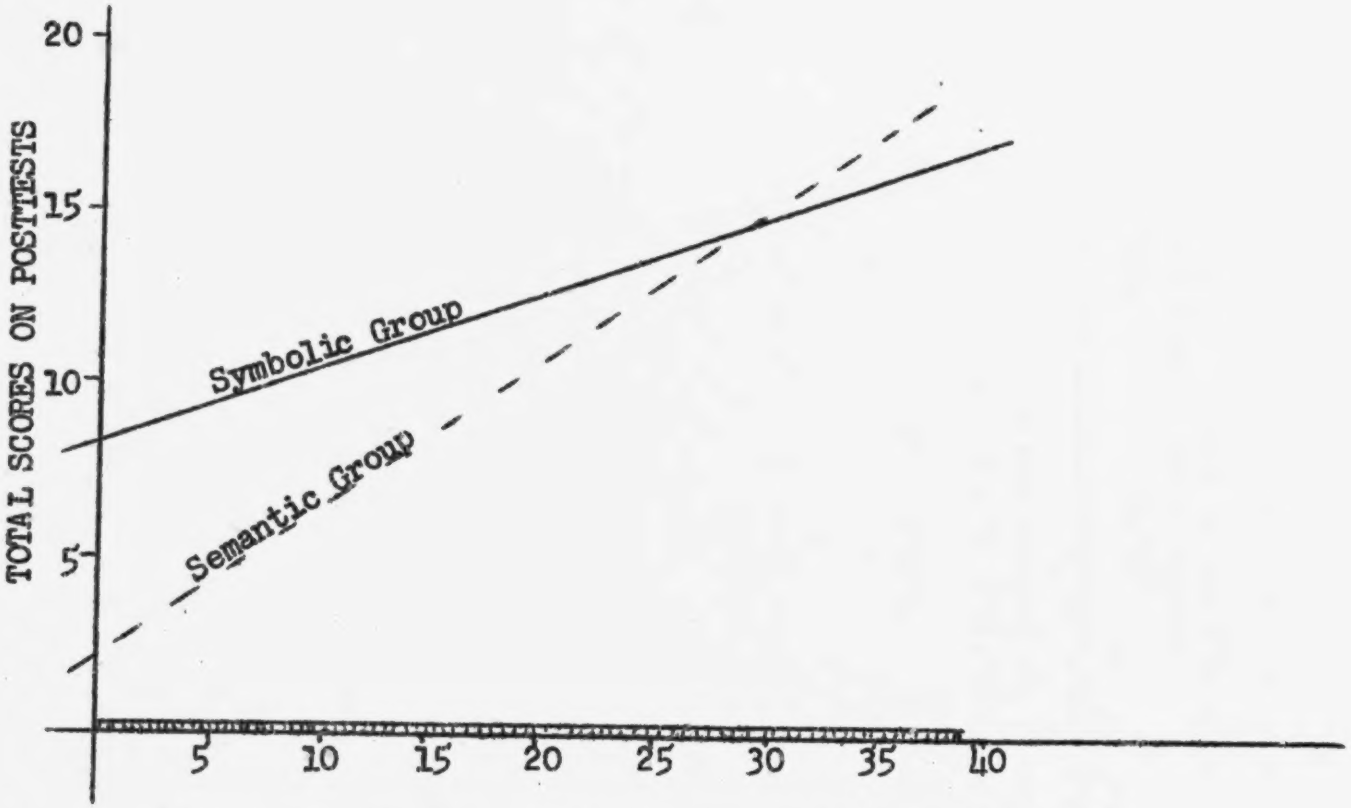

Figure 18. Graphs of simple regression equations for independent variable composed of NSR test scores.

TABLE 104

Multiple Regression Equations of Total Posttests on Transformation Tests

Group	Intercept	(NST) Regression Coefficients	(NMT) Regression Coefficients	(NST) t	(NMT) t
Semantic	4.87	.49	.68	1.60	2.41
Symbolic	6.67	.13	.92	.53	3.67

TABLE 105

Multiple Regression Equations of Vector Tests on Transformation Tests

Group	Intercept	(NST) Regression Coefficients	(NMT) Regression Coefficients	(NST) t	(NMT) t
Semantic	2.66	.27	.30	1.56	1.93
Symbolic	6.57	-.03	.29	- .20	2.25

TABLE 106

Multiple Regression Equations of Derivative Tests on Transformation Tests

Group	Intercept	(NST) Regression Coefficients	(NMT) Regression Coefficients	(NST) t	(NMT) t
Semantic	2.11	.17	.34	1.14	2.43
Symbolic	.15	.14	.52	.98	3.71

TABLE 107

Multiple Regression Equations of Derivative/Vector Tests on Transformation Tests

Group	Intercept	(NST) Regression Coefficients	(NMT) Regression Coefficients	(NST) t	(NMT) t
Semantic	.01	.08	.04	1.94	.89
Symbolic	-.09	.02	.11	.44	2.41

t-ratios are also shown for the individual multiple regression coefficients for each group relative to both factors. A graph of the multiple regression equations associated with NST and NMT test scores for a dependent variable composed of total scores on the posttest appears in Figure 19.

Means and standard deviations of the aptitude tests scores for both the semantic and symbolic groups appear in Table 108. Intercorrelations of the aptitude tests for the semantic and symbolic groups appear in Tables 109 and 110.

Correlations between ability and achievement tests and total time appear in Table 111. Intercorrelations of achievement posttest and correlations between aptitude test scores and achievement posttest scores for the semantic and symbolic groups appear in Tables 112 and 113.

Discussion

The results of this study varied considerably from those of the first two studies. Of the cognition tests, only CMR showed the predicted difference between simple regression coefficients and also corroborated the earlier studies. Two convergent production factors, NMR and NST, agreed with the two previous studies. NST coefficients did not conform to ATI theory in any of

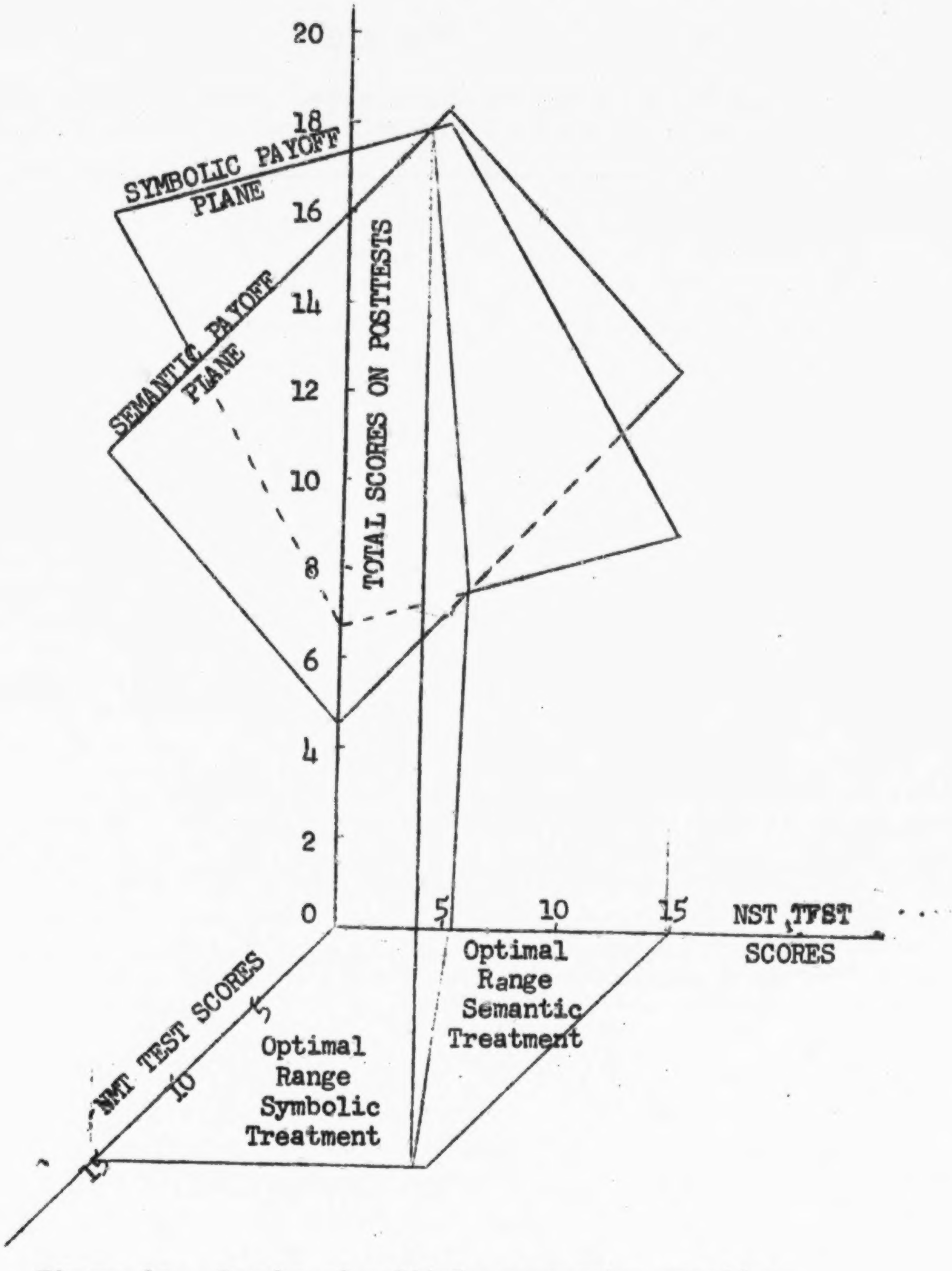

Figure 19. Graphs of multiple regression equations with independent variables composed of NST and NMT test scores.

TABLE 108

Means and Standard Deviations of Aptitude Test Scores for the Semantic and Symbolic Groups

Factor	Semantic Group (N=91) M	Semantic Group (N=91) SD	Symbolic Group (N=86) M	Symbolic Group (N=86) SD
CSR	7.68	3.76	6.71	3.37
NMR	47.76	11.97	47.12	12.07
NSR	21.98	11.44	18.52	11.85
NST	5.34	2.76	5.21	2.97
NSI	32.90	5.67	31.86	6.27
NMI	2.95	2.08	2.81	1.79
CMR	12.88	3.82	11.73	4.13
CSC	14.49	5.18	14.09	5.16
CMC	10.53	2.64	10.08	2.54
NMT	7.20	3.01	6.19	2.97

TABLE 109

Intercorrelations of Aptitude Tests

Semantic Group

TEST	CSR	NMR	NSR	NST	NSI	NMI	CMR	CSC	CMC	NMT
CSR		.46	.53	.39	.31	.37	.33	.59	.29	.26
NMR			.38	.36	.24	.30	.33	.44	.32	.28
NSR				.42	.37	.36	.47	.65	.22	.22
NST					.33	.18	.22	.30	.28	.19
NSI						.20	.27	.31	.28	.24
NMI							.30	.42	.18	.26
CMR								.51	.33	.36
CSC									.27	.26
CMC										.26
NMT										

TABLE 110

Intercorrelations of Aptitude Tests

Symbolic Group

TEST	CSR	NMR	NSR	NST	NSI	NMI	CMR	CSC	CMC	NMT
CSR		.46	.58	.47	.23	.36	.36	.53	.34	.35
NMR			.41	.30	.19	.20	.33	.36	.28	.38
NSR				.50	.25	.20	.50	.47	.42	.46
NST					.25	.23	.27	.16	.33	.31
NSI						.22	.30	.27	.36	.20
NMI							.30	.30	.27	.10
CMR								.37	.31	.44
CSC									.31	.42
CMC										.27
NMT										

TABLE 111

Correlations Between Ability and Achievement Tests and Total Time

Test	Time on Learning Program Semantic Group (N=91)	Time on Learning Program Symbolic Group (N=86)
Derivative	.30	.16
Vector	.31	.09
Derivative/ Vector	.15	.08
Total	.34	.15
CSR	.16	.01
NMR	-.08	-.28
NSR	.25	-.15
NST	.11	-.02
NSI	.31	-.12
NMI	.13	.03
CMR	.14	.03
CSC	.15	.03
CMC	.06	-.15
NMT	-.03	-.14

TABLE 112

Correlations Between Aptitude Tests and Achievement Posttests

Semantic Group

Test	Derivative	Vector	Derivative/ Vector	Total
CSR	.42	.42	.48	.49
NMR	.40	.35	.36	.43
NSR	.50	.61	.47	.64
NST	.17	.20	.22	.21
NSI	.30	.32	.15	.34
NMI	.42	.28	.41	.41
CMR	.46	.35	.27	.45
CSC	.60	.40	.50	.58
CMC	.16	.25	.22	.24
NMT	.27	.23	.13	.28
Derivative		.58	.49	.87
Vector			.52	.90
Derivative/ Vector				.65

TABLE 113

Correlations Between Aptitude Tests and Achievement Posttests

Symbolic Group

Test	Derivative	Vector	Derivative/ Vector	Total
CSR	.47	.30	.29	.46
NMR	.31	.10	.20	.26
NSR	.42	.21	.25	.38
NST	.22	.06	.13	.18
NSI	.13	.05	.13	.12
NMI	.30	.07	.31	.26
CMR	.40	.05	.22	.29
CSC	.43	.29	.33	.44
CMC	.36	.28	.34	.39
NMT	.42	.25	.28	.40
Derivative		.40	.64	.88
Vector			.39	.78
Derivative/ Vector				.73
Total				

the studies. In the multiple regression equations, the coefficient for NMT was always higher than that for NST regardless of the treatment group.

Apparently ATI effects are sensitive to differences between group educational level and/or group homogeneity. It might be that the factor tests measure different abilities at the high school and college levels. The large differences between the means of the college high school group on the criterion measures might indicate that the materials are not satisfactory for use at the high school level in the study of ATI.

D. SUMMARY

The two major purposes of the studies reported in this section were: to determine whether factor tests of semantic content would be better predictors of achievement of mathematical operations materials presented in semantic form than they would be of the same materials presented in symbolic form and whether tests of symbolic content would be better predictors of achievement of symbolic materials than they would be of semantic materials; and to determine the stability and generalizability of ATI effects.

Three studies were conducted. The first two used college freshmen as subjects. The sample sizes were 71 and 73. Subjects for the third study were 177, 10th grade high school students.

Materials and tests for all studies were the same. Two sets of learning materials were designed to teach the computation of the derivative of an algebraic expression and the multiplication of vectors. The two sets were written to be as different as possible in semantic and symbolic content. Five pairs of ability tests, differing only in semantic-symbolic content, taken from the SI model, were used as predictor variables.

In the studies involving college students, the differences between simple regression coefficients were generally stable and consistent with the ATI theory; i.e., a semantic test was usually a better predictor of achievement by students in the semantic treatment than in the symbolic treatment. The reverse was generally true for the tests of the symbolic factors.

Crossover points for a number of tests were found to be within the range of scores. This implies that achievement of the subject matter could be maximized by differentially assigning students to treatment group on the basis of their ability test scores.

Differences between multiple regression coefficients were generally less stable and conformed less well to the ATI theory.

Many of the results of the study involving high school students were at variance with those of the first two studies and with the ATI theory. Possible reasons for these discrepancies are the lower educational level and greater heterogeneity of the high school students.

X. DISCUSSION

Each study reported in the preceding sections was followed by a discussion of the specific results in terms of the expected results. Each of the preceding sections dealt with a particular area of investigation; e.g., methods of teaching vocabulary, concept formation, etc. Consequently, rather than to summarize the numerous discussions which appeared earlier, this section is devoted to a general discussion of overall methodology and problems which is applicable to all studies which were undertaken. The topics which are presented subsequently identify, what the investigators believe to be, trouble spots which should be considered before the next series of studies is formulated.

A. Interaction of Treatment-Ability-Criterion

One unanticipated difficulty encountered in several of the studies of ATI was the construction of criterion tests that would be appropriate for all treatment groups and adequately assess the common objective of the equivalent sets of materials. An example of this follows and is taken from the mathematics operations studies. The criterion items for both groups were substantively the same but they were presented in the same form as the instructional materials. If it is claimed that both groups should be able to deal with items in symbolic form, then it is obvious that the semantic group did not reach that objective. In order to reach that objective, the subjects in the semantic group would have to learn the content in verbal form first, then learn the symbol meanings, and then translate their knowledge of mathematical operations into symbolic form during their performance on the criterion. Under these conditions, it is doubtful that the mean criterion performance of the two groups would have been the same. Whether this view of the objective is reasonable depends on determining whether knowledge acquired through the two treatments will transfer equally well to the learning of subsequent material. If equal transfer occurs, then it might be concluded that the equal performance of the two groups on their respective criterion measures implied equal attainment of the objective. If equal transfer does not occur, then special attention should be given to the translation process to determine whether its facilitation could be heightened to preserve the importance of the ATI effect. If it cannot be, then one must conclude that the symbolic treatment is generally the better method of instruction. These points are of great importance in practical application of ATI theory because criterion performance in naturally occurring situations usually cannot be altered. Carroll (9) anticipated these difficulties in his suggestion that comparable scores on a criterion test might not be completely indicative of a common goal having been attained. Persons who have been taught by methods that emphasize different abilities and who have achieved comparable criterion scores might differ radically in the way they can use what has been learned in subsequent instructional settings. In addition, treatments that depend on certain abilities might serve to make persons subjected to them more highly differentiated with regard to those abilities and thus predispose them to differential achievement of subsequent instructional goals.

B. Treatment Duration

The treatment durations used in the studies, with a few exceptions, were generally set to be not greater than 50 minutes. In fact the durations were usually variable over students, so the 50-minute limit established the maximum time available to the slowest or most persistent students. This limit might be rationalized on the basis that the treatment had to fit in a class period; however, it is just as much a reflection of the great difficulty which the investigators had in preparing experimental instructional materials of any significant length at all. The brevity of these materials is undesirable on two counts. First, it might be too short, in itself and in combination with a similarly short criterion test, to permit maximum ATI effects. Second, it does not approximate even the shortest time at which students might be grouped for specialized instruction on meaningful school content. The implication of the latter statement is that although significant ATI effects were identified, it remains to be determined whether they can be identified in a typical instructional situation.

The practical significance of ATI will probably have to be established on treatment durations of at least ten hours, which parenthetically, will also establish whether specialized instructional materials can be written for an instructional duration that long.

C. Reliability of Differences

None of the studies was carried to the point at which subjects would be differentially assigned to instructional groups on the basis of identified ATI effects. Therefore, there was no occasion in the body of the report to mention or demonstrate that the success of such a practical usage would be highly dependent on the reliability of the difference scores between the aptitude measures which would be used for assignment. The reliability of the difference score is a function of the correlation between the aptitude measures and the reliability of each of them. Many of the pairs of aptitude measures which were used in the reported studies were correlated and their reliabilities were less than one might desire. When these conditions prevail, a relatively large difference between scores is needed in order to regard them as being different. The larger the difference score which is needed, then the smaller the proportion of students who might meaningfully be assigned to different treatments. The consideration of the reliability of a difference between ability scores points to another difficulty in practical application of the results of ATI studies involving two treatments and two abilities. If the too predictors have an average reliability of .90 and intercorrelation of .25, then approximately fifty percent of the persons tested will have reliably different scores on the two predictors. It would be possible to assign half of this group to each treatment, but neither treatment would be best for the remaining fifty percent of the group. This suggests that in practical situations a third treatment must be designed to care for this group. On the other hand, this difficulty is also present in studies which involve one ability or one composite and two treatments. The standard error of measurement of the ability or composite must be taken into account for scores which fall near the crossover point or plane of the regression equations. Individuals whose scores do not differ from the score at the crossover point by two standard errors probably would not profit from differential assignment. Again, it is possible that an undifferentiated or neutral treatment would be best from these individuals.

In short, the value of practical application of assignment based on ATI effects was not investigated in these studies. Subsequent studies must deal frontally with the problem.

D. Time as a Criterion

In the majority of studies, subjects worked with the instructional materials until they had finished with them, there was no fixed time limit imposed on the treatment, and then they were administered a performance test. Only in a few studies was the time at which each pupil completed the instructional materials incorporated in the analysis. In retrospect it seems that more attention should have been paid to treatment time. It is conceivable that a subject who is deficient in the aptitude which is emphasized in the instructional materials might compensate for this deficiency by spending extra time with the materials thereby causing his criterion score to be no different from that of a subject having high aptitude for the instructional materials and who completed them in far less time. Some laboratory studies might be undertaken which would have fixed and short exposures to the materials. The investigators believe the ATI effect would be sharper under that condition, and that as exposure time becomes longer and variable over subjects then it is less likely that the ATI effect will be noted. The fourth study in the series on vocabulary learning suggested, although data had not been completely analyzed, that relationships between aptitudes and performance changed throughout a period of extended instruction. It is an open question whether this might be due to development of learning sets by students, different complexities of tasks at different points in the course of instruction--which seems unlikely--or to a compensation phenomenon such as that described above.

E. Identification of Relevant Abilities

The investigators found that instructional materials which were specially devised to emphasize certain abilities and to minimize others yielded more predictable results than did studies which were executed on existing materials which were rigorously analyzed to determine what abilities they seemed to require. After consideration of this apparent differential effectiveness of the two kinds of material, the investigators grew sensitive to the possibility that the act of emphasizing a particular ability in a set of instructional materials might not necessarily result in facilitating achievement of a student who is also high on that ability. The investigators have virtually no empirical evidence on this point but do regard careful investigation of this possibility as a feasible next step in empirical studies of ATI. A geometry teacher having high verbal fluency might be coped with most satisfactorially by a student with high decoding ability or spatial visualization. A particular method of oral foreign language instruction, intended to emphasize verbal fluency, might be highly related to the student's verbal memory ability. Possibilities such as these merit investigation. Although no direct evidence was found about these kinds of interactions, the persuasiveness of the ATI theory and the failure of many of the studies to reveal interactions of the straightforward kind, e.g., materials are saturated with verbal ability consequently students with high verbal ability will do best with them, tend jointly to direct additional study to the relationship between an ability which is emphasized in a teaching act or instructional material and the ability which will be needed to cope best with it.

F. Need for Better Aptitude Measures

The value of laboratory studies of ATI and the educational application of the results of them are highly dependent on the availability of pertinent and technically excellent aptitude measures. With regard to aptitude variables, the progress of the studies was from subtests of the Primary Mental Abilities battery to the factor tests which are described by Guilford's Structure of Intellect model. The former were abandoned because they appeared to be too coarse. The latter were heavily used because a relatively large number of them was available and the SI model seemed to be a good tool with which to analyze materials to determine what aptitude factors they appeared to emphasize. However, many difficulties were encountered. First, analyses of the same materials by different people often led to the identification of different variables. Although this might be regarded as a potential ATI effect in itself, it is more likely in many instances due to vague definitions of factors by the model. Second, sometimes these identifications were of factors for which tests had not been constructed. Therefore, substitutions of similar tests had to be made and there was no way to determine the extent to which they masked ATI effects which were present and would have been revealed clearly by the most appropriate tests. Third, the aptitude measures were not orthogonal and the reliabilities were in many cases less than desirable for ATI-type research. The consequence of this is noted elsewhere in this section. In summary, penetrating studies of ATI that will lead to capitalizing on the phenomenon in practical situations will undoubtedly be hampered, perhaps prevented, by the current lack of highly suitable testing instruments.

G. Representativeness of Samples

The results of most of the studies reported herein probably somewhat underestimate the strength of ATI effects which were identified because of the nature of the samples which were used. The samples had to be drawn from available populations and these were usually located in public schools and colleges. Many of the school populations consisted of acutely disadvantaged students. The college populations undoubtedly consisted of students of superior ability. Despite the absence of norms for most of the ability measures which were used, the investigators believe that the distribution of scores of most of the samples occupied either the range at the top or the bottom of a distribution of scores which would be produced by a national sampling of people of the same age. In short, the ability ranges which were studied probably are highly restricted. Had samples been formed on the basis of a national sampling of a particular age-group, then the range would have been substantially greater, and, as a consequence, the ATI effects might have appeared when they did not and would have been much more striking than they were when they did appear. Thus, the generality of the results is somewhat limited. Generalizations to college students could probably be made with confidence, and with much less confidence to public school students due to the unrepresentativeness of most groups which were studied. Generalizations of results to educational situations typically found in the armed services and industry could be made with little confidence at all, although the investigators believe that in those situations much more striking ATI effects could be identified.

H. Generality over Age and Grade

Several studies which were replicated on groups which differed in age and educational level from the original groups did not confirm the results of the original studies. It is difficult to identify factors with confidence which might have caused this lack of uniformity of outcome. However, two factors are worthy of investigation in this regard. First, the specially prepared instructional materials might have been excellent for one of the age- or educational-level groups but not for others which were exposed to them. The difficulty level might have been appropriate for one group, but might have been too difficult or too simple for other groups, due to training or maturity, thereby diminishing or eliminating the sought for interactions. Second, because no normative data were available for most of the ability measures, one could not determine the mean score and the variability of ability scores for particular age or educational groups nor the extent of change or constancy of these characteristics over time. Both average performance and variability of performance would have direct bearing on whether ATI effects would appear in different age or educational groups when all were studied with a constant set of instructional materials.

XI. CONCLUSIONS, IMPLICATIONS AND RECOMMENDATIONS

A. Conclusions

The general purpose of the project, which incorporated the studies reported herein, was to explore the feasibility of research on ATI theory and its potential practical implications for education. The results of these studies in the broadest sense certainly demonstrate the feasibility and practicability of studies of ATI phenomena and furnish many insights about possible applications to instructional materials and procedures. The investigators regard the results of the project as strongly suggesting the wisdom of further research, theoretical and practical, and the possible practical payoff of such additional research.

Because of the exploratory nature of most of the studies which were undertaken and the fact that they were deliberately aimed to deal with different contents and processes, particular conclusions about the studies must be quite tentative. The following conclusions are ones for which a significant evidential case can be made.

(1) ATI effects exist in a variety of subject matter contents. The present studies identified ATI in mathematics learning, vocabulary learning, reading and chemistry achievement.

(2) ATI effects occur at a number of age and grade levels. The present studies found evidence for ATI in the intermediate grades, high school, and college.

(3) ATI effects are relatively replicable but they do not necessarily generalize to age and grade levels different from those of the original population.

(4) The Structure of Intellect Model is useful in conceptualizing ATI effects and in task analysis procedures. Future studies of ATI might lead to modifications of the model. The present studies seem to provide evidence of construct validity for several SI factors.

B. Implications and Recommendations

The major implication of these studies is that the investigation of ATI is a highly promising field of research with potentially significant practical applications. It seems reasonable to think that achievement of students can be enhanced by assigning them to instructional materials known to be optimally related to their ability patterns. The present studies examined only a limited number of abilities and treatments but it is conceivable that in the future computer assisted instruction facilities will be able to tailor a specific curriculum for each student by taking into account his learning history and his ability pattern.

The specific implications and recommendations which follow should be regarded as additions to or further developments of those which were presented earlier in the Chapter X, Discussion. It dealt with methodological and other problems which arose in the conduct of the studies. These problems ought to be taken into account before formulating further studies and many of them might serve as foci of research in their own right.

The following recommendations for future research in ATI grow out of the problems encountered in the present research and its positive results:

(1) Future investigations should be concerned with determining the stability of ATI effects over extended periods of time. If interactions occur only in the early periods of instruction, then their practical implications for classroom learning would be diminished. If on the other hand they persisted, become more intense, or changed with time, then their practical value might be demonstrated.

(2) The ability of students to transfer learning acquired through the use of specialized

instructional materials should be studied to mastery of subsequent material or to job performance. If poor transfer results, then the gain in learning due to specialized instruction might be nullified.

(3) Future research should pay special attention to developing more adequate procedures of task analysis. This might be accomplished by more detailed instruction of persons serving as analysts and/or by modifying the theoretical bases of the analysis.

(4) Modification or special construction of aptitude measures to increase their reliability and differential validity should be attempted. It might be achieved by lengthening existing SI measures or changing them from subject-matter content-free instruments to tests of processes within the subject matter of concern.

(5) The studies reported here were concerned only with cognitive aptitudes. Other studies should investigate ATI effects with non-cognitive variables.

(6) Future research should determine whether treatments that depend on certain abilities serve to make persons who are subjected to them more highly differentiated with regard to those abilities. If so, the question of whether this would result in differential attainment of subsequent goals should also be investigated.

(7) Research should be conducted to determine whether aptitude patterns can be altered to the extent that all students can optimally receive one treatment. It might be that this approach would lead more easily to optimal attainment of common educational goals than would the approach taken in the present project.

XII. SUMMARY

The two major purposes of the study were as follows: (a) to identify form of content variables which exist in instructional materials and which might inhibit or facilitate achievement of these materials by students of different ability patterns; and (b) to construct or identify equivalent sets of instructional materials (treatments) that differ in level of one or more form of content variables and to determine empirically the existence of aptitude treatment interactions (ATI) in their achievement by students who vary on relevant ability measures.

Seven series of studies were conducted in pursuit of the two purposes: the first series attempted to identify form of content variables in existing textual materials. Graduate students read selected passages of textbooks and recorded features of them that might facilitate or inhibit learning. In addition they ranked a set of ability measures in terms of their importance in the achievement of the material in the passages. The results showed a variety of features which might have inhibiting or facilitating effects on learning. The majority of these features were highly particular and did not appear to provide a good basis for ATI investigations.

The second series of studies dealt with redundancy as a form of content variable. The first two studies used high school students as subjects in relating elements of style to redundancy levels and in identifying cognitive abilities related to redundancy. It was felt that these studies would provide information that would allow the preparation of equivalent sets of material that differed in redundancy levels and the aptitudes which would produce ATI effects. The third and fourth studies used a set of graded-reading material with fifth- and sixth-grade students and demonstrated that they differed in redundancy level. ATI effects were shown for reasoning ability and redundancy levels.

The third series investigated ATI effects in four parallel sets of materials designed to teach elementary set concepts to fifth- and sixth-grade children. The sets of material were written to emphasize the following ability combinations: verbal-deductive, verbal-inductive, figural-deductive, and figural-inductive. Multiple regression equations were computed separately for each of four groups of students who received the four sets of material. Two criterion measures, verbal and figural, and six appropriate ability measures were included in the analyses. The patterns of regression coefficients indicated possible ATI effects but they did not conform to the hypothesized patterns. The four groups were combined into verbal and figural groups and then recombined into deductive and inductive groups, and multiple regression equations were computed using only ability measures relevant to the groupings as predictors. The results for the verbal-figural groups showed no ATI effects but the results for the deductive-inductive grouping conformed to the expectations of the ATI theory.

The purpose of the fourth investigation was to determine whether achievement of students of CHEM-STUDY is related to similarity of teacher and pupil aptitude patterns when general ability is held constant. Seven relatively pure cognitive ability measures were used as independent variables and a combination of unit and midterm achievement tests of CHEM-STUDY was the criterion. Coefficients of pattern similarity were computed for each of 13 teachers and his students. The male and female students were divided into three pattern similarity groups and these were subdivided into two general ability level groups. The results of a covariance analysis indicated that students similar in patterns to their teachers had greater achievement than did students with less similar patterns. Simple regression coefficients for each ability and achievement measure were computed separately for the students of each of six teachers. No set of regression coefficients for a particular ability appeared to be related to the teachers' scores for that variable. The results support the ATI theory prediction of a relationship between pattern similarity and achievement. The importance of a one-to-one similarity in student-teacher abilities could not be demonstrated.

The fifth series of studies investigated ATI effects in a standard concept formation task in which the instances were presented either verbally or figurally. Ability measures were selected as a result of a task analysis based on Guilford's Structure of Intellect (SI) model. Criterion measures were the number of instances selected and the number of hypotheses offered. It was hypothesized that figural concept learning would be related to two figural ability factors, CFU and CFC and that verbal concept mastery would be related to a semantic factor CMC and a symbolic factor CSU. Four concept tasks were presented by computer assisted instruction facilities or human experimenter. A pilot study indicated that ATI effects might be present in the results but a major study did not corroborate these findings. Failure to support the ATI theory might have been due to the unreliability of the criterion measures, unfamiliarity of the students with CAI equipment, or to an inadequate task analysis resulting in inappropriate ability measures.

The purpose of the sixth series of studies was to investigate ATI effects in vocabulary learning. The objective of the first study was to identify abilities that might interact with two methods of learning the meanings of difficult words. Each of two groups of graduate students who were familiar with the SI model worked through one set of materials and the synonym matching criterion test and then ranked sixteen semantic abilities from the SI model in terms of their importance in learning the materials. There appeared to be sufficient within group agreement and between group disparity in the rankings to warrant using the ability measures in ATI studies. The second and third studies were a pilot and major study designed to investigate empirically differential relationships between (a) a set of eight ability measures, which were identified in the first study, and a divergent production test and (b) vocabulary learning by three different methods. Two of the methods were used in the first study. High school students served as subjects for the pilot study and college students served as subjects for the major study. The results generally supported the ATI theory but they did not support the raters' perceptions of the order of importance of the abilities. The fourth study dealt with changes in ATI over an extended period of time.

The subjects were two groups of high school students. The results suggest that ATI effects are modified over the course of instruction. However, this generalization is highly tentative because of small, dissimilar samples of subjects, and because the data were not completely analyzed by the time this report was prepared.

The purposes of the seventh series of studies were as follows: (a) to determine whether factor tests of semantic content would be better predictors of achievement of mathematical operations presented in semantic form than they would be of the same materials presented in symbolic form, and whether tests of symbolic content would be better predictors of achievement of symbolic materials than they would be of semantic materials; and (b) to determine the stability and generalizability of ATI effects. Three studies were conducted. The first two used college freshmen as subjects and the third used tenth-grade high school students as subjects. Materials and tests for all studies were the same. Two sets of materials, one emphasizing semantic abilities and the other emphasizing symbolic abilities, were constructed. Five pairs of ability tests differing only in semantic-symbolic content were taken from the SI model and used as predictors. In the studies involving college students, the results showed generally stable ATI effects that usually conformed to the theoretical expectations. Many of the results of the study based on high school students were different from the results of the first two studies and from the ATI theory predictions. However, evidence of ATI effects was present.

A number of problems involved in studying ATI effects were identified. These were concerned with the difficulties inherent in constructing criterion measures that reflect common educational goals, problems of transfer of learning based on specialized material to subsequent learning or performance, modification of ATI effects as a function of treatment duration, lack of adequate ability measures, development of an adequate conceptual framework for task analyses, need for representative samples in future studies, and the lack of generalizability of ATI effects over different age- and grade-levels.

It was concluded that ATI effects are present in a variety of subject matter areas, at several age- and grade-levels, and appear to be replicable. A general conclusion was that the results of this exploratory project demonstrate the feasibility of attempting to make practical applications of ATI concepts to classroom situations.

XIII. REFERENCES

1. Allen, M.S., Guilford, J.P., and Merrifield, P.R. The Evaluation of Selected Intellectual Factors by Creative Research Scientists. Reports from the Psychological Laboratory of the University of Southern California, No. 25, April 1960. 11 p.

2. Beard, J.G. The Relation of Redundancy to Content Area and Expressional Style in High School Textbooks. Unpublished doctoral dissertation. Florida State University, April 1966. 74 p.

3. Biological Sciences Curriculum Study. Biological Science: Molecules to Man. Blue Version. Boston: Houghton Mifflin Company. 1963. 840 p.

4. Blade, M.F., and Watson, W.S. "Increase in Spatial Visualization Test Scores During Engineering Study," Psychological Monographs. No. 397, 69, 1955. 13 p.

5. Bond, G.L., and Cuddy, M.C. Stories to Remember: Classmate Edition. Chicago: Lyons and Carnahan. 1956. 419 p.

6. Bond, G.L., and Cuddy, M.C. Stories to Remember. Chicago: Lyons and Carnahan. 1956.

7. Bruner, J.S.; Goodnow, J.J.; and Austin, G.A. A Study of Thinking. New York: John Wiley & Sons. 1956. 330 p.

8. Bunderson, C. Victor. Transfer Functions and Learning Curves: The Use of Ability Constructs in the Study of Human Learning. Research Bulletin 64-62. Educational Testing Service. 1964. 47 p.

9. Carroll, John B. "Instructional Methods and Individual Differences," Discussion of

Dr. Cronbach's paper. In Gagne', Robert M. (ed.), *Learning and Individual Differences. Columbus: Charles E. Merrill Books, Inc.* 1967. p. 40-44.

10. Cleare, B. E. *An Investigation of the Interaction Between Student-Teacher Cognitive Ability Patterns Using Achievement in Chemical Education Material Study Chemistry Course as the Criterion Variable*. Unpublished doctoral dissertation, Florida State University. August, 1966. 64 p.

11. *Cooperative School and College Ability Tests: Level 3*. Cooperative Test Division. Educational Testing Service. 1956.

12. Cronbach, L. J., and Gleser, G. C. *Psychological Tests and Personnel Decisions. Second Edition. Urbana: University of Illinois* Press. 1965. 347 p.

13. Cronbach, L. J. "How Can Instruction be Adapted to Individual Differences," In Gagne', Robert M. (ed.), *Learning and Individual Differences. Columbus: Charles Merrill Books, Inc. 1967*. p. 23-39.

14. Dunham, J. L.; Guilford, J. P.; and Hoepfner, Ralph. *Abilities Pertaining to Classes and the Learning of Concepts. Reports from the Psychological Laboratory* of the University of Southern California, No. 39. 1966. 39 p.

15. Dunn-Rankin, Peter. *Range of Rank Totals and Its Application to Psychological Scaling. Unpublished doctoral dissertation, Florida* State University. August, 1965. 105 p.

16. Ferguson, G. A. "On Learning and Human Ability," *Canadian Journal of Psychology*. 8, 1954. *p. 95-114*.

17. Ferguson, G. A. "On Transfer and the Abilities of Man," *Canadian Journal of Psychology*. 10, 1956. *p. 121-131*.

18. Fitzpatrick, F. L; Bain, T. D.; and Teter, H. E. *Living Things*. New York: Holt, Rinehart and *Winston, Inc.* 1962. 420 p.

19. Fleishman, E.A. "A Factor Analysis of Intra-Task Performance on Two Psychomotor Tests," Psychometrika. 18, 1953. p. 45-55.

20. Fleishman, E.A., and Hempel, W. "Changes in Factor Structure of a Complex Psychomotor Task as a Function of Practice," Psychometrika. 19, 1954. p. 239-252.

21. Fleishman, E.A., and Hempel, W. "The Relation Between Abilities and Improvement with Practice in a Visual Discrimination Task," Journal of Experimental Psychology. 49, 1955. p. 301-312.

22. French, J.W.; Ekstrom, R.B.; and Price, L.A. Manual for Kit of Reference Tests for Cognitive Factors. Princeton: Educational Testing Service. 1963. 123 p.

23. Gagné, R.M. "Ability Differences in the Learning of Concepts Governing Directed Numbers," in Research Problems in Mathematics Education, OE-12008, Cooperative Research Monograph No. 3; U.S. Government Printing Office. 1960. p. 112-133.

24. Gagné, R.M. "Implications of Some Doctrines of Mathematics Teaching for Research in Human Learning," in Research Problems in Mathematics Education, OE-12008, Cooperative Research Monograph No. 3; U.S. Government Printing Office. 1960. p. 49-56.

25. Gagné, R.M., and Paradise, N.E. "Abilities and Learning Sets in Knowledge Acquisition," Psychological Monographs. No. 518, 75, 1961. p. 1-23.

26. Gagné, R.M. (ed.) Learning and Individual Differences. Columbus: Charles Merrill Books, Inc. 1967. 265 p.

27. Getzels, J.W., and Jackson, P.W. "The Teacher's Personality and Characteristics," in Gage, N.L. (ed.) Handbook of Research on Teaching. Chicago: Rand McNally and Co. 1963. p. 506-582.

28. Guilford, J.P., and Hoepfner, R. Structure of Intellect Factors and Their Tests. Reports

from the Psychological Laboratory of the University of Southern California. No. 36. 1966. 15 p.

29. Haggard, E.A. Intraclass Correlation and the Analysis of Variance. New York: The Dryden Press. 1958. 171 p.

30. Harlow, H.F. "The Formation of Learning Sets," Psychological Review. 56, 1949. p. 51-65.

31. Johnson, D.M., and Stratton, R.P. "Evaluation of Five Methods of Teaching Concepts,: Journal of Educational Psychology. No. 1, 57, 1966. p. 48-53.

32. Kenney, J.F., and Keeping, E.S. Mathematics of Statistics: Part Two. Princeton: D. van Nostrand. 1951. 429 p.

33. Kohfeld, D.L. "The Prediction of Perceptual-Motor Learning from Independent Verbal and Motor Measures," Psychonomic Science. 4, 1966. p. 413-414.

34. Kohler, E.T. An Investigation of Cloze Scores in Terms of Selected Cognitive Variables. Unpublished doctoral dissertation, Florida State University. 1966.

35. Kroeber, E.; Wolff, W.H.; Weaver, R. L. Biology, Second Edition. Boston: D. C. Heath and Co. 1965. 646 p.

36. Kropp, R.P., and Stoker, H.W. The Construction and Validation of Tests of the Cognitive Processes as Described in the Taxonomy of Educational Objectives. Unpublished report, Cooperative Research Project No. 2117, Office of Education, U.S. Department of Health, Education, and Welfare. February 1966. 434 p.

37. Peterson, H.; Guilford, J.P.; Hoefpner, R.; and Merrifield, P.R. Determination of "structure-of-intellect" Abilities Involved In Ninth-Grade Algebra and General Mathematics. Reports from the Psychological Laboratory of the University of Southern California. No. 31. 1963. 40 p.

38. Ramsey, J.G. The Attainment of Concepts from Figural and Verbal Instances by Individual and Pairs. Technical report No. 4. Research and Development Center on Learning and Re-education, University of Wisconsin. 1965. 13 p.

39. Snow, R.E.; Tiffin, J.; and Siebert, W.F. "Individual Differences and Instructional Film Effects," Journal of Educational Psychology. 56, 1965. p. 315-326.

40. Sullivan, E.T.; Clark, W.W.; and Tiegs, E.W. California Test of Mental Maturity: Level 2. California Test Bureau. 1963.

41. Textbooks in Print 1965. New York: R.R. Bowker Co.,1965. 420 p.

42. Thorndike, E.L., and Lorge, I.L. The Teacher's Workbook of 30,000 Words. New York: Bureau of Publications, Teachers College. 1944. 274 p.

43. Thurstone, T.G. Primary Mental Abilities: Grades 4-6. Science Research Associates, Inc. 1962.

44. Wallen, N.E., and Travers, R.M.W. "Analysis and Investigation of Teaching Methods," in Gage, N.L. (ed.) Handbook of Research on Teaching. Chicago: Rand McNally & Co. 1963. p. 448-505.

45. Winer, B.J. Statistical Principles in Experimental Design. New York: McGraw-Hill Book Co., Inc. 1962. 672 p.